AN INVESTIGATION INTO THE PROPERTIES OF BAYESIAN FORECASTING MODELS

by

Nick S. Cantarelis

A thesis submitted for the degree of Doctor of Philosophy

School of Industrial and Business Studies,
University of Warwick,
Coventry CV4 7AL

December 1979

TABLE OF CONTENTS

Page

Page

Page

CHAPTER 5 : UNDERSTANDING AND CONTROLLING THE RESPONSE OF THE SYSTEM

Page

CHAPTER 6 : ON LINE ESTIMATION OF V_ε IN THE MSM

CHAPTER 7 : ESTIMATION OF VARIANCES IN THE STEADY STATE DLM

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I am grateful to Dr. F.R.Johnston for his numerous invaluable suggestions, advice and encouragement in every stage of this research. I also wish to thank the C.I.E.B.R. , the Umbrako fund and the S.I.B.S. for their contribution towards the costs of the work. Finally I wish to thank the members of the Computer Unit staff and in particular Keith Halstead, Tim Clark, Gerry Sheridan and Ruth Raymond.

SUMMARY

In the early 70's, Harrison and Stevens [17, 18, 19], made a major contribution to the area of statistical forecasting. They adopted a Bayesian approach in conjunction with a fundamental model, first used by Kalman [24] and called the Dynamic Linear Model (DLM).

This thesis is concerned with the investigation of the properties of different Bayesian forecasting models and in particular of the Multi State Model (MSM). The latter is important in postulating that no single DLM can adequately describe a process with discontinuities and consequently the system defines a number of models characterising the most likely process states. A small number of parameters are shown to govern the behaviour of the MSM and the relationship between the choice of these parameters and performance has been examined. It is shown that for a process exhibiting discontinuities, traditional forecasting criteria such as the mean square error are no longer appropriate. An alternative set of performance measures is proposed and used as the main language of understanding the variety of responses of the MSM to different types and sizes of discontinuities. The parameter representing the noise variance of the process is shown to be critical to the performance of both the MSM and other single state Bayesian models. A number of on line variance estimation methods are proposed and tested on artificial and real data. The methods are shown to be robust and lead to improved performance not only of the MSM but the other Bayesian single state models which of course require a noise variance estimate.

Finally, alternative formulations of the MSM are proposed, leading to significant reduction in the computational and storage requirements while at the same time improving the response of the MSM.

To

My parents and Maria

CHAPTER 1

INTRODUCTION

1.1. Definition of time series

The analysis of time series is an important area of statistics with application in a variety of fields such as economics, engineering, production planning, quality control and many others.

A time series is defined as a collection of observations y_1 y_2 .... y_n made sequentially in time and the analysis of time series is essentially concerned with identifying the stochastic model which might have generated these observations.

Khintchine [26,27], showed that a time series may be viewed *as a random process of the variable* Y_t *sampled over time periods* $t = 1, 2, \ldots, n$. Hence y_t for $t = 1, 2 \ldots n$ may be thought of as a particular realisation of an infinite set of time series which might have been observed by sampling from Y_t.

Similarly Kolmogorov [23] viewed it as a sample from an n-dimensional probability distribution with each time period corresponding to one dimension of this distribution.

A major objective of time series analysis is to predict future values of the series and this is an important task in sales forecasting which is the main area of interest of this thesis.

1.2. The role of short term sales forecasting in management

Mathematical methods for producing routine short term sales forecasts have slowly become accepted as an important source of information for effective decision making.

In inventory control systems forecasts of customer demand are needed each week or month to be used for stock control, ordering raw materials and packages, production planning etc.

The United States National Industrial Conference Board reporting its study [38] on Business Policies in Inventory Management states that "nearly 30% of the working capital of the average business is tied up in inventories".

Lewis [31] reports that the interest lost on money invested in stock can add approximately 20% to the works prime cost.

If improved methods of producing forecasts can reduce the uncertainties of future demand then lower stock holding, scheduling and production planning can be achieved.

1.3. Historical development of time series methods

Historically the theory of time series developed from Statistics. It therefore became concerned with stationary time series, i.e. where the mean or level of the series is constant and the variability retains the same pattern through time. A more mathematical definition can also be given using the notation of section 1.1 :

A time series is stationary if the joint probability distribution of Y_{t+1} , Y_{t+2} , ... Y_{t+n} is the same with that of Y_{t+k+1} , Y_{t+k+2} , ... Y_{t+k+n} for all t, n and k.

Real data series are usually non-stationary but by suitable transformations or differencing stationarity can sometimes be achieved.

The interest in time series started in 1927 with the work of Yule [46] who introduced the concept of autoregression. His approach was further developed in 1931 by Walker [39] who defined the general autoregressive (AR) process of (1.3.1). This assumes that an observation y_t can be represented as the sum of a weighted linear combination of old observations and some chance element:

$$y_t = \sum_{i=1}^{p} \phi_i \, y_{t-i} + \varepsilon_t \qquad (1.3.1)$$

where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ for all t.

In 1937 Slutsky [36] presented the moving average (MA) process of (1.3.2.):

$$y_t = \sum_{j=1}^{q} \Theta_j \, \varepsilon_{t-j} + \varepsilon_t + \mu \qquad (1.3.2)$$

where μ is the constant mean of the process.

A major advance in the theory took place in 1938 when Wold [44] developed a comprehensive theory of mixed autoregressive/ moving averages (ARMA) processes. He proved that any discrete stationary process can be expressed as the sum of two uncorrelated processes, one purely deterministic (e.g. a sinusoidal process) and one purely indeterministic (e.g. MA and AR processes).

The first complete solution to the least squares prediction problem was given in 1941 by Kolmogorov [29] for discrete stationary time series. The problem was to predict a future value y_{t+k} using a linear predictor $\hat{y}_{t+k}$ where

$$\hat{y}_{t+k} = \sum_j c_j \, y_{t-j}$$

such that $E(y_{t+k} - \hat{y}_{t+k})^2$ is a minimum.

Mann and Wold [33] in 1943 solved the problem of estimating the unknown parameters in stationary autoregressive series using maximum likelihood methods and Whittle [41], Durbin [12] and Walker [40] extended their results to more general processes such as ARMA and multiple time series.

During World War II Wiener [49] working in the field of control and communication pioneered the estimation problem for continuous filters (the continuous counterpart of Kolmogorov's least squares prediction problem). He developed new techniques for filtering a signal at the receiver whose transmission had been distorted by some random noise process.

Wiener showed that the prediction of random signals and their separation from random noise can be achieved by solving the so-called Wiener-Hopf integral equation and he gave a method for its solution in the practically important special case of stationary processes.

Before 1960 many extensions and generalisations followed Wiener's basic work but all of them were subject to a number of limitations which seriously curtailed their practical usefulness. In 1960-61 however, Kalman and Bucy [24, 25], proposed a new approach to the standard filtering and prediction problem for multivariate non stationary Markov processes. They realised that rather than to attack the Wiener-Hopf integral equation directly, it is better to convert it into a non linear differential equation whose solution yields the covariance matrix of the minimum filtering error, which in turn contains all necessary information for the design of the optimal filter.

The result of their work was the Kalman filter recursive equations, which are easily updated on line on receipt of additional observations.

Details of the Kalman Filter will be given later on since it turns out to be a fundamental feature of Bayesian forecasting.

Exponential smoothing models were introduced around 1960 by operational researchers such as Brown [6,7], Holt [22] and Winters [43]. They are special cases of autoregressive models with the weights placed on past observations decreasing exponentially. Their main contribution has been in short term sales forecasting and stock control.

In the late 60's Box and Jenkins [3, 5] made an important contribution to the theory of time series forecasting. Their procedure utilises the concepts of MA, AR, ARMA processes but its major advantage is that a class of models is proposed together with a strategy by which for any given series a particular model is chosen from this class according to the properties of the particular time series in question. Thus as Newbold [35] puts it, the form of the eventual forecast function is dictated, to a large extent by the data - a principle known as "letting the data speak for itself."

In contrast to the exponential smoothing models the Box-Jenkins procedure is not fully automatic in the sense that given a data series, a computer program can not be used to produce forecasts without manual intervention, since a single model has to be chosen from the large class of autoregressive integrated moving average (ARIMA) models.

Very briefly an ARIMA (p, d, q) process is defined as follows:

Consider a process $\{y_t \; ; \; t = 1, 2 \ldots.\}$ which can be non stationary. Then it is assumed that y_t can be reduced to stationarity by successive differencing and that an integer d exists such that

$$w_t = (1 - B)^d \, y_t$$

is stationary, where $B^j y_t = y_{t-j}$

It is further assumed that w_t can be represented as an ARMA model:

$$w_t = \sum_{i=1}^{p} \phi_i \, w_{t-i} + \sum_{j=1}^{p} \Theta_j \, \varepsilon_{t-j} + \varepsilon_t \qquad (1.3.3)$$

and this is known as an ARIMA (p, d q) model where:

ϕ_i are the autoregressive parameters

Θ_j are the moving average parameters

and ε_t is a white noise sequence distributed as $N(0, \sigma_\varepsilon^2)$

Box and Jenkins (B/J) fit ARIMA models to a set of data by an iterative three stage cycle of identification, estimation and diagnostic checking. First tentative values of p,d, q are chosen which allows the estimation of ϕ_i and Θ_j for i = 1, 2 ... p and j = 1, 2 ... q. Finally, the adequacy of the fitted model is checked and any inadequacies found are expected to suggest alternative values for p, d, q and hence the cycle is repeated until a satisfactory model obtains.

Clearly the B/J procedure is more versatile than exponential smoothing models many of which are simply subsets of the general class considered by B/J. However, Chatfield [48] argues that in practice some exponential smoothing models are not special cases of the B/J procedure and points out that the multiplicative Holt-Winters forecasting procedure does not have an ARIMA equivalent. B/J methods are of course much more expensive and considerable experience is required to identify an appropriate ARIMA model. Often (e.g. see Chatfield [11]) several such models may be found to fit the data equally well but can produce different forecasts. Recently considerable ingenuity has gone into removing the subjective element in the first stage of the cycle (i.e. identification of p, d, q) assuming a statinnary process. Akaike [2] is the main name associated with this and in engineering circles a well known approach, State Space Forecasting as presented by Mehra [34] for example, is equivalent to an automatic technique for selecting p, d, q. This simplifies the B/J procedure for the user without denying its subtlety but the major objection is its reliance on relative stability in the process which in socioeconomic time series is certainly quite unrealistic. Finally a serious drawback of the B/J procedure is the requirement of a moderately long series of data for its satisfactory application.

In the early 70's almost parallel to the work of B/J, Harrison and Stevens (H/S) [17, 18, 19], developed a Bayesian approach pioneering the use of Kalman filters in time series forecasting. Following the filtering results of Kalman they showed how previously accepted techniques can be restructured using a Markov formulation of time series. This structural change covers all existing linear time series models and with the introduction of likelihood methods the Bayesian approach provides a means of learning and communication. The advantages offered by this approach are numerous and an attempt is made in the following section to summarise them.

1.4. The case for Bayesian forecasting

In order to appreciate fully the facilities and advantages offered by the Bayesian approach it is first necessary to introduce the Dynamic Linear Model (DLM) which is a special case of the state space characterisation of linear systems developed by Kalman. The general form of the univariate DLM is characterised by two equations:

Observation eq$^{\underline{n}}$: $y_t = \underline{F}_t \underline{\theta}_t + \varepsilon_t$; $\varepsilon_t \sim N(0, V_t)$ (1.4.1)

System eq$^{\underline{n}}$: $\underline{\theta}_t = \underline{G}\ \underline{\theta}_{t-1} + \underline{w}_t$; $\underline{w}_t \sim N(\underline{0}, \underline{W}_t)$ (1.4.2)

where y_t is the process observation made at time t

$\underline{\theta}_t$ is an $(n \times 1)$ unobservable vector of process parameters at time t.

$\underline{F}_t$ is a $(1 \times n)$ vector of independent variables known at time t and defines how the process is observed.

$\underline{G}$ is an $(n \times n)$ known system matrix whose function is to represent deterministic changes in $\underline{\theta}_t$ from one time period to the next.

ε_t is a stochastic noise element representing errors inherent in observing the true $\underline{\theta}_t$.

$\underline{w}_t$ is an $(n \times 1)$ stochastic disturbance vector representing random shocks during the evolution of $\underline{\theta}_t$.

$V_t = E(\varepsilon_t^2)$ assumed known at all times

$W_t = E(\underline{w}_t\ \underline{w}_t')$ assumed known at all times

Equation (1.4.1) specifies how y_t, the process observation at time t, is stochastically dependent on the current process parameter vector $\underline{\Theta}_t$ while equation (1.4.2) specifies how the process parameters evolve in time, both stochastically and deterministically through $\underline{w}_t$ and $\underline{G}$ respectively.

It follows from (1.4.2) that $\underline{\Theta}_t$ evolves according to a stochastic process in the presence of random shocks and hence it can never be known exactly. Thus at time t the information about $\underline{\Theta}_t$ is expressed in terms of a normal distribution:

$$(\underline{\Theta}_t \mid D_t) \sim N(\underline{m}_t, \underline{C}_t)$$

where D_t is the set of all the data up to and including time t, i.e.

$$D_t = \{D_o, y_1, \underline{F}_1, y_2, \underline{F}_2 \ldots\ldots y_t, \underline{F}_t\} = \{D_{t-1}, y_t, \underline{F}_t\}$$

where D_o represents the information supplied at the start prior to any observations e.g. $\underline{m}_o$ and $\underline{C}_o$ expressing our views on $\underline{\Theta}_t$ at time $t = 0$.

Given an additional observation $(y_{t+1}, \underline{F}_{t+1})$, $\underline{\Theta}_t$ is updated using the Kalman filter recurrence relations which will be given in the next chapter.

We can now discuss a number of advantages of the Bayesian approach in an attempt to illustrate its importance in the forecasting field.

a) Communication

This is a consequence of the parametric structural representation of time series by the DLM.

Consider for example the following deterministic linear growth process y_t for $t = 1, 2 \ldots$:

$$y_t - 2y_{t-1} + y_{t-2} = 0 \qquad (1.4.3)$$

The DLM representation of this process is:

$$\left.\begin{aligned} y_t &= \mu_t \\ \mu_t &= \mu_{t-1} + \beta_t \\ \beta_t &= \beta_{t-1} \end{aligned}\right\} \qquad (1.4.4)$$

where the parameters μ_t , β_t are easily interpreted as the "level" of the process at time t and the incremental "growth" in level in the interval t - 1 to t.

Hence the Markovian structure of the DLM allows an easier interpretation of the model parameters with the result that subjective knowledge is more easily assimilated into the scheme. In this respect the Bayesian approach has a distinct advantage over ARIMA models. The forecaster can now impart his views on the parameters reflecting information external to the data history. This is not only useful initially when no data history exists but it is also important in anticipating and dealing with major process changes. The point is that an environmental change which invalidates the whole representation usually only invalidates one part or a limited number of parts and often does not affect the overall model concept which in our previous example is captured by both equations (1.4.3) and the system (1.4.4). The difference lies in the fact that if the overall model fails at time t because we think that for some reason the level of the process has changed but the growth is the same, then this subjective information can easily be communicated to the DLM by changing the current estimate of the level without disturbing the growth estimate. At this point in time it is natural that our uncertainty has also increased as a result of this environmental change and this can also be communicated by simply increasing the variance estimates associated with the process level and possibly of growth too.

A critical distinction is made by H/S [18] between a statistical *forecasting method* and a *forecasting system*. The former transforms input data into output information in a purely mechanical way while the latter includes people; the person responsible for the forecasts as well as all the people concerned with using the forecasts and supplying information relevant to the resulting actions.

b) Recursiveness

A major weakness of all traditional forecasting methods has been their inability to deal adequately with a new time series such as occur at product launch or even when there is a major change in the market for a product.

The recursive nature of the Kalman filter however implies that the current posterior distribution

$$(\underline{\theta}_t \mid D_t) \sim N(\underline{m}_t, \underline{C}_t)$$

may be calculated from

(i) the most recent observation, y_t

(ii) the current observation and disturbance variance V_t and $\underline{W}_t$

and (iii) the posterior distribution

$$(\underline{\theta}_{t-1} \mid D_{t-1}) \sim N(\underline{m}_{t-1}, \underline{C}_{t-1})$$

The great advantage of this is that, for the first time, we have a rational, coherent framework for forecasting in situations where there is little or no prior data history and for developing systems which require structured models to facilitate communication between man and "routine forecasting method".

c) Uncertainty about the underlying model

Another disadvantage of traditional forecasting methods is their inability to reflect varying uncertainty not only with regard to parametric estimates but also with a range of possible models and their complete absence of any form of model learning.

The Bayesian approach however is geared towards learning, change and probability descriptions and provides elegant ways of extending the DLM, thus overcoming these problems.

A DLM is characterised by $\underline{F}_t$, $\underline{G}$, V_t and $\underline{W}_t$ so that a particular model $M^{(j)}$ can be specified as:

$$M^{(j)} \equiv (\underline{F}_t^{(j)} , \underline{G}^{(j)} , V_t^{(j)} , \underline{W}_t^{(j)})$$

If a number of such models is considered for j = 1, 2 ... n with associated initial prior probabilities $p_o^{(j)}$ then Bayes' theorem can be used to update these probabilities with every new observation, leading to an on line model identification procedure.

H/S [18] term this model set "multi process model" and distinguish between two possibilities:

(I) Class I multi process model:
Here it is assumed that a single unknown model adequately represents y_t and either

the learning logic of Bayes' theorem is used to select the single most appropriate model for the process under observation,

or

forecasts and decisions are based on a weighted combination of the whole set of models so that at any time the selected model is not necessarily one of the original models but an "interpolated" model.

(II) Class II multi process model:

The postulate here is that at different periods of time, different models or combinations of models appropriately describe the context. In other words the process shifts from being best represented by one model to being best represented by another.

H/S [19] describe a situation with four different process states namely "no change", "outlier", "growth change", and "step change". These states are characterised by different variances $V_t^{(j)}$, $\underline{W}_t^{(j)}$ in each of four linear growth stochastic generating models $M^{(j)}$ of the form:

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t^{(j)} & \varepsilon_t^{(j)} &\sim N\,(0,\ V_\varepsilon^{(j)}) \\
\mu_t &= \mu_{t-1} + \beta_t + \delta\mu_t^{(j)} & \delta\mu_t^{(j)} &\sim N\,(0,\ V_\mu^{(j)}) \\
\beta_t &= \beta_{t-1} + \delta\beta_t^{(j)} & \delta\beta_t^{(j)} &\sim N\,(0,\ V_\beta^{(j)})
\end{aligned}
$$

where $M^{(j)}$ for j = 1, 2, 3, 4 is used to model the four different process states. The variance magnitudes necessary to characterise these states are shown below and further explanation and justification for these will be given in Chapter 3.

TABLE 1.1.

j	STATE j	$V_\varepsilon^{(j)}$	$V_\mu^{(j)}$	$V_\beta^{(j)}$
1	"no change"	Normal	Nil	Nil
2	"outlier"	Large	Nil	Nil
3	"growth change"	Normal	Nil	Large
4	"step change"	Normal	Large	Nil

Allowing models to continually wax and wane in a probabilistic fashion necessitates statements on model transition probabilities and an obvious choice is a very high probability for the "no change" model and low probabilities for the "intruding" models.

This system has the advantage of recognising and responding appropriately to sudden changes of level, slope and outliers which are major problems for all previous forecasting methods. Because of the different states modelled by this multi process Class II model, it will often be referred to as *multi state model* (MSM).

The learning capability of this approach is obvious. Because of the Bayesian foundations of the DLM it is possible to start forecasting immediately at time zero when the forecasts will be based entirely on the initial prior probabilities of parameters and models. However, as time progresses the probabilities are continually modified by the data and hence there is no clear distinction, as in conventional methods, between start up, model selection, parameter estimation and actual forecasting, since this is done on line.

d) Decision making

Traditional forecasting methods are not capable of communicating varying degrees of uncertainty. Yet in practice this is always the case: at product launch or when a sudden change has taken place there is a lot of uncertainty which decreases as data becomes available and a stable pattern is established.

This changing uncertainty is automatically reflected in a MSM for example, through the changing posterior probabilities of the various models which are combined to produce realistic measures of the forecast error variance for the purpose of decision making. In addition probabilistic information on the process parameters at any given time means that decisions are not based on a single figure forecast but a probability distribution.

e) Stationarity

Virtually all previous forecasting methods were based on the assumption that the observed series is stationary or can be reduced to stationarity by a suitable transformation and can then be used for model identification. This assumption is however too strong since, in identification, this implies that because something never happened in the past it will never happen in the future which is not realistic. In the Bayesian approach there is no stationarity requirement since the evolution of the future is not wholly tied to that of past observations nor to a single model.

This list of advantages although not exhaustive gives sufficient information for the potentials of the Bayesian approach to be fully recognised.

1.5. Objectives and structure of the thesis

The prime objective of this thesis has been to investigate the properties and examine methods of overcoming a number of theoretical and practical difficulties which arise in the application of Bayesian models. Most of the work however is concentrated on the MSM since it is undoubtably the most important model relieving the forecaster of the burden of model identification while being computationally cheaper than B/J procedures for example, since it does not require any non linear search routines.

The following topics are examined:

(i) Definition of forecasting performance: Chapter 4 looks at the problem of measuring the performance of a forecasting system since the traditional mean square error (MSE) criterion is shown to be inappropriate and misleading. This is especially obvious when the MSM is used to model data series exhibiting a number of discontinuities. Instead a mixture of both qualitative and quantitative measures are proposed and justified by a number of arguments and empirical evidence.

(ii) Sensitivity and robustness of the MSM : At time period zero the MSM requires information to be supplied in terms of a number of parameters. Some of these parameter values are updated by the system whilst others are not.

The objective of Chapter 5 is to determine the sensitivity and robustness of the MSM in relation to the choice of those parameters which remain fixed throughout since their choice controls the response of the system.

(iii) Estimation of variance : One problem of the Bayesian approach is as follows. The Kalman filter assumes an exact knowledge of the noise and disturbance variances which must be known outside the system. This is a crucial problem to the successful practical application of the approach but very little attention has been paid to it in the literature and especially to the important case of on line variance estimation. Godolphin and Harrison [15] recognise this explicitly pointing out that parameter search can give much difficulty to the inexperienced worker using the DLM Markov representation since it takes him into the relatively unfamiliar territory of variance rather than coefficient estimation.

The problem arises because if the models operate with fixed estimates of these variances then the predictive distributional information available from the Kalman filter (which are very important for decision making) can be misleading. In addition an estimate of the "Normal" variability as defined earlier in Table 1.1 is of special importance to the MSM which in essence interpretes deviations from forecast in the light of prediction error. If for example this "Normal" variability is overestimated then the prediction error will be similarly overestimated and the system will respond slower to a genuine sudden change in the environment. This is because the larger forecast errors resulting from this change will initially be interpreted as "no change" in the light of our over-estimated expected forecast errors.

The objective of Chapters 6 and 7 is to investigate this variance problem and devise methods for its on line estimation.

(iv) Computational efficiency of the MSM and its speed of response to growth changes: As pointed out earlier it is true that in relative terms the MSM is computationally cheap. However as will be seen in Chapter 2 it is possible to model the effects of seasonality, promotional activity etc. by incorporating a suitable number of parameters into the process parameter vector $\underline{\theta}_t$. Suppose for example that $\underline{\theta}_t$ consists of the usual level and growth components as well as 13 seasonal factors and possibly 13 promotional ones. Then the MSM operates with vectors and square matrices of order 28. Under such circumstances as well as when the MSM is to be used for forecasting a large number of items, the computational and storage requirements may well be too high for many potential users of the system. With this in mind Chapter 8 looks at ways of making the MSM more computationally efficient by excluding a number of unlikely state transitions.

Another objective in that Chapter was to investigate alternative formulations of the MSM in order to improve its performance and particularly its speed of response to abrupt changes in growth. The problem of growth response improvement was one of great interest to a number of practitioners who were starting to experiment with the MSM but had difficulties in "tuning" the system so as to respond faster to growth changes without at the same time becoming over-sensitive during "quiet" periods.

Chapter 9 summarises our conclusions and a number of proofs, graphs and computer program listings are given in the Appendices.

Finally, Chapters 2 and 3 describe some fundamental concepts of the Bayesian approach thus completing the necessary background.

1.6. NOTATION AND ABBREVIATIONS

The notation will be carefully defined in each Chapter and for all the abbreviations used the term will be given in full at the first instance followed by the abbreviation in brackets. A list of symbols and abbreviations together with the section when they were first used or defined is given in Appendix A and can be used as a general reference whenever the meaning of a symbol or term is ambiguous.

Subsection z of section y of chapter x will be referred to as x. y. z. and the n^{th} equation of x. y. z will be referred to as (x. y. z. n).

The k^{th} reference will be denoted by [k].

Finally, "Table x. y" refers to the y^{th} table which can be found in chapter x and similarly for "Figure x. y ".

CHAPTER 2

THE BAYESIAN APPROACH TO FORECASTING

2.1. Foreword

The aim of this chapter is to summarise some basic theoretical results of the Bayesian approach. The *Kalman filter* parameter estimation and updating procedure is first described and the *principle of superposition* is introduced which has important implications in model building. This is followed by two sections giving the DLM representation and properties of two widely used models. These will be used for reference from several parts of the thesis.

2.2. The Kalman filter

Recall the general form of the DLM as defined in section 1.4:

Observation eq$^{\underline{n}}$: $y_t = \underline{F}_t \ \underline{\theta}_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, V_t)$ (2.2.1)

System eq$^{\underline{n}}$: $\underline{\theta}_t = \underline{G} \ \underline{\theta}_{t-1} + \underline{w}_t \qquad \underline{w}_t \sim N(0, \underline{W}_t)$ (2.2.2)

Suppose that at time t-1 our posterior information about the process parameter vector $\underline{\theta}_t$ takes the form of a Normal distribution,

$$(\underline{\theta}_{t-1} \mid D_{t-1}) \sim N(\underline{m}_{t-1}, \underline{C}_{t-1}) \qquad (2.2.3)$$

then as soon as y_t and $\underline{F}_t$ become available the estimation problem is to make inferences about the current value of the unknown parameter vector $\underline{\Theta}_t$. This is exactly what the Kalman filter does and it can be shown (see Appendix B) that as a consequence of the DLM representation and the distribution of $(\underline{\Theta}_{t-1} \mid D_{t-1})$, the posterior distribution of $\underline{\Theta}_t$ given D_t is also Normal,

$$(\underline{\Theta}_t \mid D_t) \sim N\,(\underline{m}_t\,,\,\underline{C}_t) \tag{2.2.4}$$

where $\underline{m}_t$ and $\underline{C}_t$ are calculated recursively from the following Kalman filter equations first given by Kalman [25].

$$\left.\begin{aligned}
\hat{y}_t &= E\,(y_t \mid D_{t-1}) = \underline{F}_t\,\underline{G}\,\underline{m}_{t-1} \\
e_t &= y_t - \hat{y}_t \\
\underline{R}_t &= \mathrm{Var}\,(\underline{\Theta}_t \mid D_{t-1}) = \underline{G}\,\underline{C}_{t-1}\,\underline{G}' + \underline{W}_t \\
\hat{Y}_t &= \mathrm{Var}\,(y_t \mid D_{t-1}) = \underline{F}_t\,\underline{R}_t\,\underline{F}'_t + V_t \\
\underline{A}_t &= (\hat{Y}_t)^{-1}\,\underline{R}_t\,\underline{F}'_t
\end{aligned}\right\} \tag{2.2.5}$$

$$\underline{m}_t = E\,(\underline{\Theta}_t \mid D_t) = \underline{G}\,\underline{m}_{t-1} + \underline{A}_t\,e_t \tag{2.2.6}$$

$$\underline{C}_t = \mathrm{Var}\,(\underline{\Theta}_t \mid D_t) = \underline{R}_t - \underline{A}_t\,\hat{Y}_t\,\underline{A}'_t \tag{2.2.7}$$

The analogy of $\underline{A}_t$ in (2.2.6) with traditional fixed smoothing constants is obvious since $\underline{A}_t$ controls the amount of the forecast error fed back into revising the parameter estimates. The difference is that $\underline{A}_t$ is now a smoothing vector which is not fixed but varies with time in a manner determined by the relative uncertainties.

Of course at time $t = 0$ prior to any observations the Kalman filter requires an initial prior distribution,

$$(\underline{\Theta}_t \mid D_0) \sim N(\underline{m}_0 \; , \; \underline{C}_0)$$

utilising information based on experience, analogy, judgement etc. This initial subjective information requirement is an important feature of the approach and it can be argued that it is a step which many practitioners are not only willing to take but do take on the many occasions in real life where decisions implying forecasts must be made prior to the occurence of the first data point.

Having established the procedure of inferring $(\underline{\Theta}_t \mid D_t)$ we now simply state the results derived by H/S [18] which can be used to make inferences about future values of the variable under observation in the important case of $\underline{F}_t$ being constant or varying in a predictable way for example on account of periodic functions. Using the observation and system equations of the DLM it can be shown that the prediction mean and variance of the k-step ahead process observation can be obtained from:

$$\left.\begin{aligned} \hat{y}_{t+k} &= \underline{F}_{t+k} \; \hat{\underline{m}}_{t+k} \\ \hat{Y}_{t+k} &= \underline{F}_{t+k} \; \hat{\underline{C}}_{t+k} \; \underline{F}'_{t+k} \; + \; V_{t+k} \end{aligned}\right\} \qquad (2.2.8)$$

where $\hat{\underline{m}}_{t+k}$ and $\hat{\underline{C}}_{t+k}$ are recursively obtained from:

$$\hat{\underline{m}}_{t+k} = E(\underline{\Theta}_{t+k} \mid D_t) = \underline{G}\, \hat{\underline{m}}_{t+k-1}$$

$$\hat{\underline{C}}_{t+k} = Var(\underline{\Theta}_{t+k} \mid D_t) = \underline{G}\, \hat{C}_{t+k-1}\, \underline{G}' + \underline{W}_{t+k}$$

for $k = 1,2,3\ldots$ When $k = 1$, $\hat{\underline{m}}_{t+k-1}$ and $\hat{\underline{C}}_{t+k-1}$ are in fact the latest Kalman filter estimates of $\underline{\Theta}_t$ and its uncertainty:

$$\hat{\underline{m}}_{t+k-1} = \hat{\underline{m}}_t = E(\underline{\Theta}_t \mid D_t) = \underline{m}_t$$

$$\hat{\underline{C}}_{t+k-1} = \hat{\underline{C}}_t = Var(\underline{\Theta}_t \mid D_t) = \underline{C}_t$$

Finally we comment on the *principle of superposition* which has far reaching implications for the construction of large sophisticated models. The principle simply states that a linear combination of a number of simpler linear models is itself a linear model. Thus the model builder might fix his attention in simple sub-systems which he can readily describe in DLM form and then bring them together into a single DLM. For example, in a sales forecasting situation the forecaster might construct a sub-model DLM describing the effect of advertising promotions, another DLM to describe the seasonal variation and finally to describe the wandering and effect of unrecorded factors as a linear growth time series DLM.

The power of the principle of superposition can be illustrated in terms of the following simple example. Suppose that we are operating with a model of the general DLM form given by (2.2.1) and (2.2.2) but we believe that the observation noise,

$$\varepsilon_t \sim N(0, V_t)$$

would be better described by a correlated time series such as an expontentially weighted moving average or ARIMA (0,1,1) process. As will be seen in the next section such a process for ε_t can be represented as follows:

$$\varepsilon_t = \phi_t + \varepsilon_t^* \qquad \varepsilon_t^* \sim N(0, V_\varepsilon^*)$$

$$\phi_t = \phi_{t-1} + \delta\phi_t \qquad \delta\phi_t \sim N(0, V_\phi)$$

Using the principle of superposition the complete model may be expressed as follows:

$$y_t = [\underline{F}_t \;,\; 1] \begin{bmatrix} \underline{\theta}_t \\ \phi_t \end{bmatrix} + \varepsilon_t^*$$

$$\begin{bmatrix} \underline{\theta}_t \\ \phi_t \end{bmatrix} = \begin{bmatrix} \underline{G} & \underline{0} \\ \underline{0} & 1 \end{bmatrix} \begin{bmatrix} \underline{\theta}_{t-1} \\ \phi_{t-1} \end{bmatrix} + \begin{bmatrix} \underline{w}_t \\ \delta\phi_t \end{bmatrix}$$

which with an obvious redefinition of $\underline{F}_t$ $\underline{\theta}_t$ etc. can be seen to be another DLM of the general form.

In sections C.1 and C.2 of Appendix C an attempt is made to specify the way in which the principle of superposition can be used to build Bayesian systems of any degree of generality. The model of C.1

includes seasonality factors in the parameter vector $\underline{\theta}_t$ while that of C.2 incorporates promotion and price effects in addition to seasonality. Both represent actual practical applications of Bayesian systems and have formed parts of M.Sc. projects prepared for two large manufacturing companies based in London and Liverpool respectively.

2.3. Properties of the constant variance Steady State Model (SSM)

The SSM is widely used in practice to model a process which can be generated by a simple random walk. In sales forecasting for example the model can be used to predict the demand for a product which is established and reasonably stable. It assumes that the observed figure of demand is the sum of some underlying true level and some unpredictable random disturbance. Further the underlying true level at any period is equal to that in the preceeding period plus some small random perturbation. Such a process can be written as:

$$y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, V_\varepsilon) \qquad (2.3.1)$$

$$\mu_t = \mu_{t-1} + \delta\mu_t \qquad \delta\mu_t \sim N(0, V_\mu) \qquad (2.3.2)$$

where μ_t = unknown true process level at time t

ε_t = observation noise at time t

$\delta\mu_t$ = level disturbance at time t

and V_ε , V_μ are unknown true constant variances of the uncorrelated random variables ε_t and $\delta\mu_t$. This process gives rise to the simplest DLM representation:

$$\left.\begin{array}{llcl} \text{Observation eq}^{\underline{n}} & : \; y_t & = & [1]\,[\mu_t] + \varepsilon_t \\ \text{System eq}^{\underline{n}} & : \; [\mu_t] & = & [1]\,[\mu_{t-1}] + [\delta\mu_t] \end{array}\right\} \qquad (2.3.3)$$

Comparing this representation with the general DLM formulation given in Section 1.4 it can be seen that for the SSM,

$$\left.\begin{array}{llll} n = 1 & \underline{\theta}_t = [m_t] & \underline{F}_t = [1] \\ \underline{G} = [1] & V_t = V_\varepsilon & \underline{W}_t = [V_\mu] \\ \text{and} & \underline{w}_t = [\delta\mu_t] & \end{array}\right\} \qquad (2.3.4)$$

At time $t = 0$ information about μ_0 is described in the form of a Normal distribution,

$$(\mu_0 \mid D_0) \sim N(m_0 \;,\; C_0)$$

and estimates for the true variances V_ε , V_μ are nominated and designated $V_{\varepsilon,N}$ and $V_{\mu,N}$. Alternatively $V_{\varepsilon,N}$ and $r_{\mu,N}$ can be nominated where $r_{\mu,N}$ is our estimate of the true process variance ratio $r_\mu = (V_\varepsilon/V_\mu)$ which can be viewed as a measure of the time dependence of μ_t. Using this notation and substituting the SSM values of $\underline{\theta}_t$, $\underline{F}_t$, $\underline{G}$, V_t and $\underline{W}_t$ into the general Kalman filter equations (2.2.5), (2.2.6) and (2.2.7) it is easy to show that the model tends to a limiting form with the limiting values of $\underline{R}_t$, $\hat{Y}_t$, $\underline{A}_t$ and $\underline{C}_t$ given by R_N, $\hat{Y}_N$, A_N and C_N respectively :

$$\left.\begin{aligned} R_N &= V_{\varepsilon,N}\ A_N \ / \ (1 - A_N) \\ \hat{Y}_N &= V_{\varepsilon,N} \ / \ (1 - A_N) \\ A_N &= \left[-1 + (1 + 4r_{\mu,N})^{\frac{1}{2}}\right] \ / \ (2r_{\mu,N}) \\ C_N &= A_N \ V_{\varepsilon,N} \end{aligned}\right\} \qquad (2.3.5)$$

The posterior distribution for the process mean is given by :

$$\left.\begin{aligned} & (\mu_t \mid D_t) \sim N(m_t \ , \ C_N) \\ \text{where} \quad & m_t = m_{t-1} + A_N e_t \\ & e_t = y_t - \hat{y}_t \\ & \hat{y}_t = m_{t-1} \end{aligned}\right\} \qquad (2.3.6)$$

The model in this limiting form is equivalent to an ARIMA (0, 1, 1) model and an exponentially weighted moving average (EWMA) model operating with a smoothing constant α equal to A_N. Hence through A_N, the nominated ratio $r_{\mu,N}$ is directly related to α and controls the up-dating of m_t. It can be seen from the expression for A_N for example that $r_{\mu,N} = 20$ corresponds to an EWMA $\alpha = 0.2$.

In the limit our estimate of the level can easily be shown to be of the following form:

$$m_t = A_N \sum_{i=0}^{\infty} (1 - A_N)^i \, y_{t-i} \qquad (2.3.7)$$

which implies that in the limit $r_{\mu,N}$ is the only factor (except the actual data) affecting the updating of the process level and consequently the limiting value of the MSE is a function of $r_{\mu,N}$ alone and does not depend on $V_{\varepsilon,N}$. It can in fact be shown (see Appendix D) that the one step ahead forecast error e_t is distributed normally with zero mean and variance given by,

$$\text{Var}\ (e_t) = E(e_t^{\,2}) = \underset{t \to \infty}{\text{limit}}\ (\text{MSE})_t = \frac{2A_N V_\varepsilon + V_\mu}{A_N\ (2-A_N)} \qquad (2.3.8)$$

$$\text{where } (\text{MSE})_t = \frac{1}{t} \sum_{i=1}^{t} e_i^{\,2}$$

This result also shows clearly that the limiting value of MSE depends only on $r_{\mu,N}$ and not on the actual value of $V_{\varepsilon,N}$.

2.4. Properties of the constant variance Linear Growth Model (LGM)

The LGM extends the SSM by the addition of a growth component β_t and is highly relevant to applications. As an extension of the SSM of (2.3.1) and (2.3.2) it can be written as follows:

$$\text{Observation eq}^{\underline{n}} : \quad y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \sim N(0,\ V_\varepsilon) \qquad (2.4.1)$$

$$\left.\begin{aligned} &\text{System eq}^{\underline{n}} : \quad \mu_t = \mu_{t-1} + \beta_t + \delta\mu_t && \delta\mu_t \sim N(0,\ V_\mu) \\ &\qquad\qquad\qquad \beta_t = \beta_{t-1} + \delta\beta_t && \delta\beta_t \sim N(0, V_\beta\ ) \end{aligned}\right\} \quad (2.4.2)$$

If the equation for β_t is substituted in the μ_t equation then the DLM representation of the LGM is obvious.

$$y_t = \begin{bmatrix} 1 , 0 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + \varepsilon_t$$

$$\begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} \delta\mu_t + \delta\beta_t \\ \delta\beta_t \end{bmatrix}$$

and by comparison to the general DLM formulation of Section 1.4 it can be seen that for the LGM we have:

$$n = 2 \quad \underline{F}_t = \begin{bmatrix} 1 , 0 \end{bmatrix} \quad \underline{\Theta}_t = \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} \qquad (2.4.3)$$

$$\underline{G} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \underline{w}_t = \begin{bmatrix} \delta\mu_t + \delta\beta_t \\ \delta\beta_t \end{bmatrix}$$

$$V_t = \mathrm{Var}(\varepsilon_t) = V_\varepsilon \quad \text{and}$$

$$W_t = E(\underline{w}_t \, \underline{w}_t{}') = \begin{bmatrix} V_\mu + V_\beta & V_\beta \\ V_\beta & V_\beta \end{bmatrix}$$

Let the information about $\underline{\theta}_{t-1}$ at time t-1 and prior to y_t be of the form:

$$\left(\begin{bmatrix}\mu_{t-1}\\ \beta_{t-1}\end{bmatrix} \middle| D_{t-1}\right) \sim N\left(\begin{bmatrix}m_{t-1}\\ b_{t-1}\end{bmatrix} ; \begin{bmatrix}c_{11,t-1} & c_{12,t-1}\\ c_{21,t-1} & c_{22,t-1}\end{bmatrix}\right) \quad (2.4.4)$$

where m_{t-1} = best estimate of the process level given D_{t-1}

b_{t-1} = " " " " " growth " "

$c_{11,t-1}$ = variance on the estimate of the level

$c_{22,t-1}$ = " " " " " " growth

$c_{12,t-1}$ = $c_{21,t-1}$ = covariance of the level and growth estimates

Substituting the LGM values of $\underline{F}_t$, $\underline{G}$, V_t and $\underline{W}_t$ into the general Kalman filter equations (2.2.5), (2.2.6) and (2.2.7) it is straightforward to arrive at the following posterior estimates, given the latest observation at time t, y_t:

$$\left(\begin{bmatrix}\mu_{t}\\ \beta_{t}\end{bmatrix} \middle| D_{t}\right) \sim N\left(\begin{bmatrix}m_{t}\\ b_{t}\end{bmatrix} ; \begin{bmatrix}c_{11,t} & c_{12,t}\\ c_{12,t} & c_{22,t}\end{bmatrix}\right) \quad (2.4.5)$$

where

$$\left.\begin{aligned}
m_t &= m_{t-1} + b_{t-1} + A_{1,t}\ e_t \\
b_t &= b_{t-1} + A_{2,t}\ e_t \\
c_{11,t} &= r_{11,t}\ (1 - A_{1,t}) \\
c_{12,t} &= c_{21,t} = r_{12,t}\ (1 - A_{1,t}) \\
c_{22,t} &= r_{22,t} - \hat{Y}_t (A_{2,t})^2
\end{aligned}\right\} \qquad (2.4.6)$$

and

$$\left.\begin{aligned}
y_t &= m_{t-1} + b_{t-1} \\
e_t &= y_t - \hat{y}_t \\
\underline{R}_t &= \begin{bmatrix} r_{11,t} & r_{12,t} \\ r_{21,t} & r_{22,t} \end{bmatrix} = \\
&= \begin{bmatrix} c_{11,t-1} + 2c_{12,t-1} + c_{22,t-1} + V_\mu + V_\beta & c_{12,t-1} + c_{22,t-1} + V_\beta \\ c_{21,t-1} + c_{22,t-1} + V_\beta & c_{22,t-1} + V_\beta \end{bmatrix} \\
\hat{Y}_t &= r_{11,t} + V_\varepsilon \\
\underline{A}_t &= \begin{bmatrix} A_{1,t} \\ A_{2,t} \end{bmatrix} = (\hat{Y}_t)^{-1} \begin{bmatrix} r_{11,t} \\ r_{12,t} \end{bmatrix}
\end{aligned}\right\} \qquad (2.4.7)$$

Given initially nominated estimates $V_{\varepsilon,N}$, $V_{\mu,N}$ and $V_{\beta,N}$ for the true but unknown variances V_{ε}, V_{μ}, V_{β} and provided there are no interventions from management then the Kalman updating pattern will approach a limiting form satisfying the following system of equations:

$$\left.\begin{aligned} V_{\varepsilon,N} &= (1 - A_{1,N})\,\hat{Y}_N \\ V_{\mu,N} &= (A_{1,N}^2 + A_{1,N}\,A_{2,N} - 2A_{2,N})\,\hat{Y}_N \\ V_{\beta,N} &= A_{2,N}^2\,\hat{Y}_N \end{aligned}\right\} \tag{2.4.8}$$

where the subscript t has been replaced by N to emphasize that $A_{1,t}$, $A_{2,t}$ and $\hat{Y}_t$ have reached a limit which depends purely on the nominated variances $V_{\varepsilon,N}$, $V_{\mu,N}$ and $V_{\beta,N}$.

In this limiting form the LGM is equivalent to the ARIMA(0,2,2) predictor of B/J [3] as well as Holt's [23] linear growth predictor, which can be written as follows:

$$\begin{aligned} m_t &= m_{t-1} + b_{t-1} + A_1 e_t \\ b_t &= b_{t-1} + A_2 e_t \end{aligned}$$

where the well known smoothing constants A_1, A_2 are analogous to $A_{1,N}$, $A_{2,N}$ in the LGM. The difference between traditional linear growth models and the LGM is that in the latter we have no direct control over $A_{1,N}$, $A_{2,N}$ but instead these are limiting values determined by the Kalman filter and depend purely on the variance ratios:

$$\left.\begin{aligned} r_{\mu,N} &= (V_{\varepsilon,N} / V_{\mu,N}) \\ \text{and} \quad r_{\beta,N} &= (V_{\mu,N} / V_{\beta,N}) \end{aligned}\right\} \qquad (2.4.9)$$

Substituting in (2.4.8) for $V_{\varepsilon,N}$ and $V_{\mu,N}$ in terms of $r_{\mu,N}$, $r_{\beta,N}$ and $V_{\beta,N}$ from (2.4.9) it follows that there is a correspondence between pairs of $(r_{\mu,N}, r_{\beta,N})$ and $(A_{1,N}, A_{2,N})$. Four examples of this are shown in Table 2.1. It has therefore been established that the role played by $r_{\mu,N}$ and $r_{\beta,N}$ in the LGM is equivalent to that played by A_1 and A_2 in conventional linear growth models.

Hypothetical Nominated variances			Variance ratio		Corresponding smoothing constants	
$V_{\varepsilon,N}$	$V_{\mu,N}$	$V_{\beta,N}$	$r_{\mu,N}$	$r_{\beta,N}$	$A_{1,N}$	$A_{2,N}$
2500	100	1	25	100	.25	.02
250000	100	1	2500	100	.06	.002
2500	100	2500	25	(1/25)	.78	.48
2500	250000	1	(1/100)	250000	.99	.0004

TABLE 2.1
Correspondence between the ratios of nominated variances and the smoothing constants

Finally an important relationship exists between the LGM and Brown's second order predictor. Brown [6, 7] proposed a model for linear growth which is also known as the method of exponentially weighted regression (EWR). Many practitioners prefer EWR to Holt's method because it can be viewed as a formalisation of the natural approach to short term sales forecasting. Generally, a salesman's approach to forecasting when faced with a graph of customer demand is to draw some curve, usually a straight line, through the observations and to derive forecasts by projecting the curve into the future. In drawing the curve he will make it fit the most recent observations the best. This is exactly what EWR does by choosing the coefficients of the curve which minimise the exponentially weighted sum of squares of the differences between the observed demand and the curve. That is, the K-step ahead forecast at time t is,

$$\hat{y}_{t+K} = m_t + Kb_t$$

where given a discount factor w, m_t and b_t are chosen to minimise,

$$\sum_{i=0}^{\infty} w^i (y_{t-i} - m_t - ib_t)^2$$

so that the final updating equations are:

$$\left.\begin{aligned} m_t &= m_{t-1} + b_{t-1} + (1 - w^2)e_t \\ b_t &= b_{t-1} \qquad\quad\; + (1 - w)^2 e_t \end{aligned}\right\} \qquad (2.4.10)$$

Comparing equations (2.4.10) with the corresponding updating equations (2.4.6) of the LGM, it is obvious that $A_{1,N}$ and $A_{2,N}$ are related to the EWR discount factor w:

$$\left.\begin{aligned} A_{1,N} &= 1 - w^2 \\ A_{2,N} &= (1 - w)^2 \end{aligned}\right\} \qquad (2.4.11)$$

and consequently given a particular discount factor w, $A_{1,N}$ and $A_{2,N}$ can be found from (2.4.11) which if substituted in (2.4.8) fixes the LGM variance ratios $r_{\mu,N} = (V_{\varepsilon,N} / V_{\mu,N})$ and $r_{\beta,N} = (V_{\mu,N} / V_{\beta,N})$. The important implication of this is that given a set of data, the search for optimal values of $r_{\mu,N}$ and $r_{\beta,N}$ becomes a one dimensional search for the optimal discount factor. Harrison [16] has shown that the increase in standard error through using the EWR values over all possible $A_{1,N}$, $A_{2,N}$ pairs (not restricted by (2.4.11) is at most 7% and with routine industrial data where $A_{1,N}$ will be less than 0.25 the increase in error is less than 1.6%.

CHAPTER 3 THE MULTI STATE MODEL (MSM)

3.1. Foreword

The aim of this Chapter is to build the necessary background for the four state linear growth MSM which is the main model under investigation in the thesis. Full explanation of its structure and the way it operates is given and in doing so the notation is introduced to be used in the subsequent chapters. Being a Bayesian model the MSM possesses all the advantages described in Section 1.4 the main one being its ability to recognise and respond appropriately to typical types of environmental shock that arise when there are sudden market changes such as occur in the form of spot orders, competitive action, economic crisis etc.

3.2. Modelling discontinuities

Consider the constant variance LGM of section 2.4:

$$y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, V_\varepsilon) \qquad (3.2.1)$$

$$\mu_t = \mu_{t-1} + \beta_t + \delta\mu_t \qquad \delta\mu_t \sim N(0, V_\mu) \qquad (3.2.2)$$

$$\beta_t = \beta_{t-1} + \delta\beta_t \qquad \delta\beta_t \sim N(0, V_\beta) \qquad (3.2.3)$$

and suppose we are interested in modelling the following fictitious sales data exhibiting a number of discontinuities.

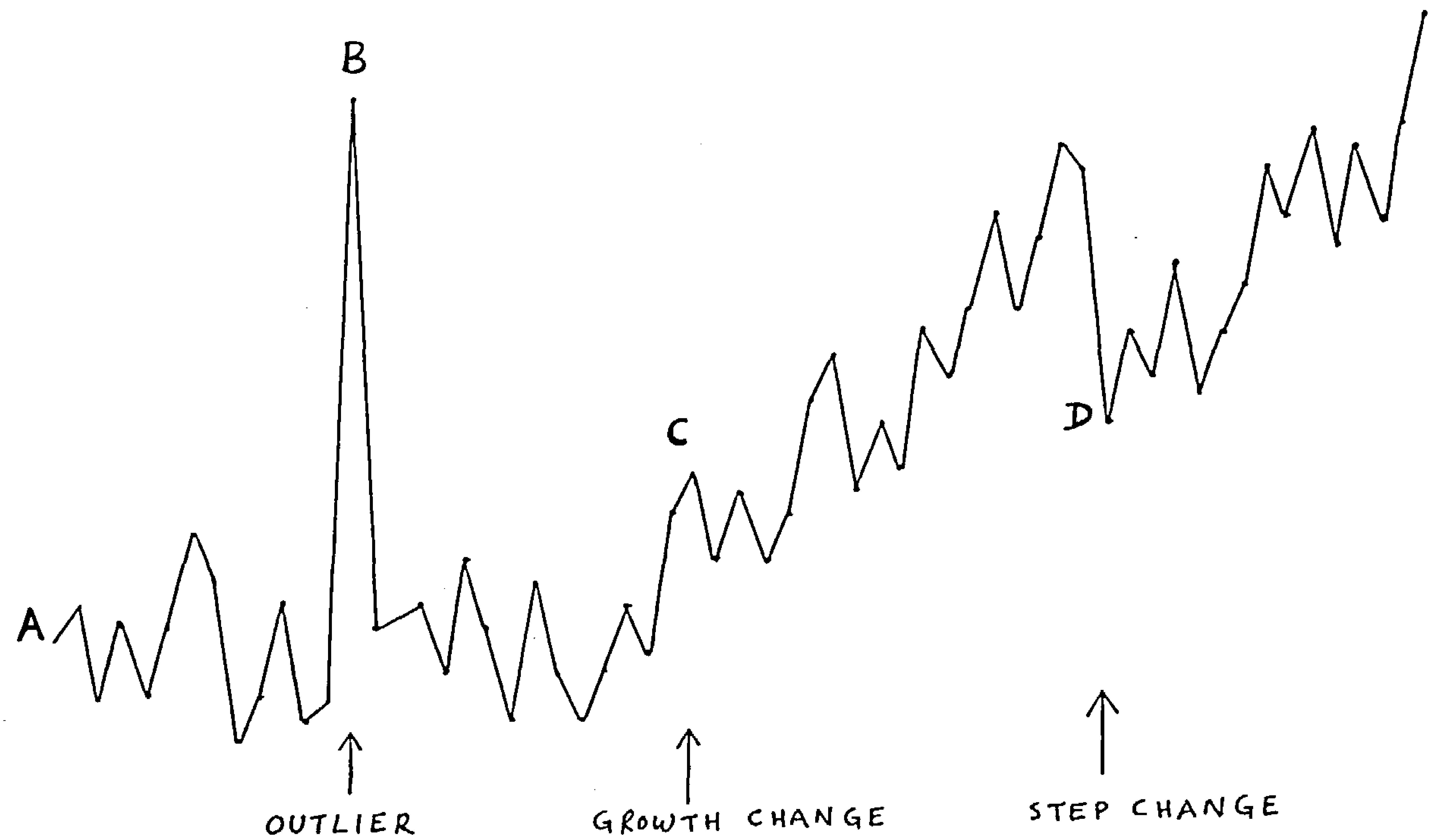


Figure 3.1

Any single model which was sensitive enough to respond fast to such abrupt changes in growth and level would be too unstable during quiet periods. It would therefore seem natural to describe the situation by several models.

The LGM as well as traditional linear growth models (e.g. Holt's, ARIMA(0,2,2), EWR etc.) are inappropriate for this type of data because they treat forecast errors in a purely mechanical way updating the current level and growth estimates by simply feeding back a fixed percentage (determined by the smoothing constants) of the forecast error, as implied by the updating equations:

$$m_t = m_{t-1} + b_{t-1} + A_1 e_t$$

$$b_t = \qquad b_{t-1} + A_2 e_t$$

where (i) m_t, b_t are the current estimates of μ_t, β_t.

(ii) e_t is the forecast error at time t.

and (iii) A_1, A_2 are the smoothing constants which are either directly specified as in conventional models or are determined by $V_{\varepsilon,N}$, $V_{\mu,N}$, $V_{\beta,N}$, as shown in Section 2.4 for the LGM.

The failure of these models at points of major discontinuities is a consequence of the fact that they treat unusually large forecast errors (such as would occur at points B, C, D, in figure 3.1) in exactly the same way as usual acceptable forecast errors. It is worth noting here that a modification proposed by Trigg and Leach [47] to forecasting systems employing exponential smoothing whereby the response rate is varied and made to depend on the value of a tracking signal leads to an improved response to step changes.

The MSM attempts to interpret forecast errors as the result of the occurrence of one out of four possible process states:

state 1 : "no change"

" 2 : "outlier"

" 3 : "growth change"

" 4 : "step change"

It is possible to model these process states by supposing that at each point in time the observation noise ε_t and the level and growth disturbances $\delta\mu_t$, $\delta\beta_t$ are generated from not one (as implied by the LGM formulation of equations (3.2.1), (3.2.2), (3.2.3) but four possible distributions:

$$\varepsilon_t \sim N(0, V_\varepsilon^{(j)}) \tag{3.2.4}$$

$$\delta\mu_t \sim N(0, V_\mu^{(j)}) \tag{3.2.5}$$

$$\delta\beta_t \sim N(0, V_\beta^{(j)}) \tag{3.2.6}$$

for $j = 1, 2, 3, 4.$

Definition : Let $M_t^{(j)}$ denote a model which at time t characterises state j by a particular combination of $V_\varepsilon^{(j)}$, $V_\mu^{(j)}$ and $V_\beta^{(j)}$ for $j = 1,2,3,4.$

Then $M_t^{(j)}$ can be thought of as putting forward a theory to explain the forecast error at time t, e_t. In the next section it will be shown how posterior probabilities $p_t^{(j)}$ can be calculated at any time t, representing the probability that $M_t^{(j)}$ is currently offering the best theory for the explanation of e_t.

To determine the relative magnitudes necessary for $V_\varepsilon^{(j)}$, $V_\mu^{(j)}$ and $V_\beta^{(j)}$ in order to model the four states we consider the following example. Suppose for simplicity that at time $t-1$ we know the values of the true process parameters and they are:

$$\mu_{t-1} = 980$$

$$\beta_{t-1} = 20$$

It follows from (3.2.1), (3.2.2), (3.2.3) that the observation at time t, y_t, can be written as

$$y_t = \mu_{t-1} + \beta_{t-1} + \varepsilon_t + \delta\mu_t + \delta\beta_t = 1000 + \varepsilon_t + \delta\mu_t + \delta\beta_t \qquad (3.2.7)$$

Hence $E(y_t) = 1000$ and therefore the forecast error can be written as follows:

$$e_t = y_t - E(y_t) = \varepsilon_t + \delta\mu_t + \delta\beta_t \qquad (3.2.8)$$

It now follows from (3.2.1), (3.2.2), (3.2.3) and (3.2.8) that the distribution of the forecast error is:

$$\left.\begin{aligned} & e_t \sim N(0, V) \\ \text{where} \quad & V = V_\varepsilon + V_\mu + V_\beta \end{aligned}\right\} \qquad (3.2.9)$$

and the distribution of e_t conditional on $M_t^{(j)}$ is then:

$$\left.\begin{array}{lll} & (e_t \mid M_t^{(j)}) \sim N(0, V^{(j)}) & \\ \text{or} & (\varepsilon_t + \delta\mu_t + \delta\beta_t \mid M_t^{(j)}) \sim N(0, V^{(j)}) & \\ \text{where} & V^{(j)} = V_\varepsilon^{(j)} + V_\mu^{(j)} + V_\beta^{(j)} & \\ \text{for} & j = 1,2,3,4 & \end{array}\right\} \quad (3.2.10)$$

We can now consider the implications of distinct combinations of $V_\varepsilon^{(j)}$, $V_\mu^{(j)}$, $V_\beta^{(j)}$ and their relative magnitudes necessary to characterise the four states.

3.2.1. Model of a "no change" state : $M_t^{(1)}$

If in the time interval $(t-1, t)$ the process has remained in a "no change" state then the implication is that,

$$\mu_t \sim \mu_{t-1} \quad \text{i.e.} \quad \delta\mu_t \sim 0$$

$$\beta_t \sim \beta_{t-1} \quad \text{i.e.} \quad \delta\beta_t \sim 0$$

and the forecast error e_t is simply the result of the observation noise ε_t. Hence the theory put forward by $M_t^{(1)}$ is that e_t has been generated by:

$$e_t \sim N(0, V^{(1)})$$

$$\text{with} \quad \left.\begin{aligned} V_\varepsilon^{(1)} &= V_\varepsilon \\ V_\mu^{(1)} &= 0 \\ V_\beta^{(1)} &= 0 \end{aligned}\right\} \qquad (3.2.1.1)$$

where V_ε is the basic acceptable variability of the process when no change is taking place. The way V_ε is selected for a particular process is explained in Section 3.4 of this chapter. The fact that the combination of variances given by (3.2.1.1) is a sensible way of modelling a "no change" state is clear since $V_\mu^{(1)} = V_\beta^{(1)} = 0$ imply (using (3.2.5) and (3.2.6)) that $\delta\mu_t = \delta\beta_t = 0$ so that the level and growth parameters μ_t, β_t are unchanged, which is the case when "no change" has taken place in the time interval $(t-1, t)$.

3.2.2. Model of an "outlier" state : $M_t^{(2)}$

Given that at time t an outlier has occurred the implication is that the forecast error e_t is unusually large. After this "wild" observation however, the process continues in an unaffected way, that is with unchanged level μ_t and growth β_t. $M_t^{(2)}$ should therefore view e_t $(= \varepsilon_t + \delta\mu_t + \delta\beta_t)$ as simply the result of an unusually large observation noise ε_t with $\delta\mu_t = \delta\beta_t = 0$ since non-zero $\delta\mu_t$, $\delta\beta_t$

would imply that μ_t, β_t have changed as a result of the outlier. Hence $M_t^{(2)}$ can model an outlier by viewing e_t as having been generated by:

$$\left.\begin{aligned} e_t &\sim N(0, V^{(2)}) \\ \text{with} \quad V_\varepsilon^{(2)} &= \lambda_2 V_\varepsilon \qquad \lambda_2 \gg 1 \\ V_\mu^{(2)} &= 0 \\ V_\beta^{(2)} &= 0 \end{aligned}\right\} \qquad (3.2.2.1)$$

The differing views of $M_t^{(1)}$ and $M_t^{(2)}$ when explaining the forecast error e_t are illustrated below.

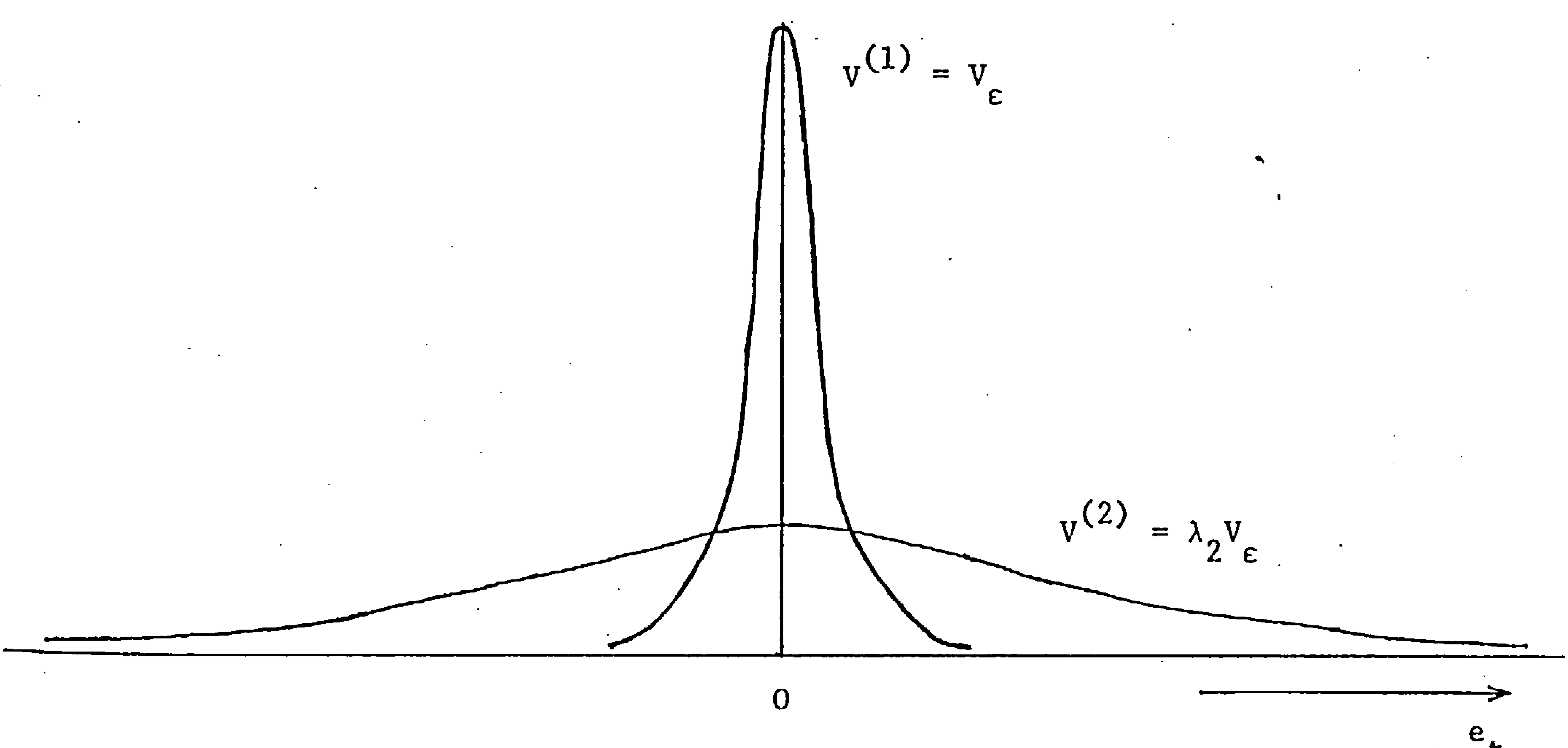


Figure 3.2

To obtain a reasonable estimate of λ_2 let us suppose that we wish the probability given by $M_t^{(2)}$ for values of $e_t \geq 10\,\sigma$ ($\sigma = \sqrt{V_\varepsilon}$) to be equal to the probability that $M_t^{(1)}$ attributes to values of $e_t \geq 1\sigma$. This is equivalent to the following probability statement:

$$p(e_t \geqslant 10\sigma \mid M_t^{(2)}) = p(e_t \geqslant 1\sigma \mid M_t^{(1)})$$

which is true when:

$$\sqrt{V^{(2)}} = 10\,\sqrt{V^{(1)}}$$

or $$\sqrt{\lambda_2 V_\varepsilon} = 10\,\sqrt{V_\varepsilon}$$

leading to $\lambda_2 = 100$.

3.2.3. Model of a "growth change" state : $M_t^{(3)}$

If an abrupt growth change has taken place in the time interval $(t-1, t)$ then $|\beta_t - \beta_{t-1}| >> 0$ which implies (from equation (3.2.3)) that $\delta\beta_t$ is unusually large. It follows that just after the growth change the forecast errors will be larger than average (that is larger than what $M_t^{(1)}$ expects them to be, e.g. 3σ or 4σ) and therefore an alternative more likely model than $M_t^{(1)}$ is necessary which views the forecast error as made up by (i) a usual observation noise ε_t contribution i.e. $\varepsilon_t \sim N(0, V_\varepsilon)$ explaining part of

e_t $(= \varepsilon_t + \delta\mu_t + \delta\beta_t)$ plus (ii) an unusually large contribution of $\delta\beta_t$, i.e. $\delta\beta_t \sim N(0, \lambda_3 V_\varepsilon)$ $\lambda_3 > 0$ with (iii) $\delta\mu_t = 0$ since a non-zero $\delta\mu_t$ would imply a "step change", which is not the case.

Hence $M_t^{(3)}$ can model a "growth change" by viewing e_t as having been generated by:

$$e_t \sim N(0, V^{(3)})$$

$$\text{with} \quad \left.\begin{aligned} V_\varepsilon^{(3)} &= V_\varepsilon \\ V_\mu^{(3)} &= 0 \\ V_\beta^{(3)} &= \lambda_3 V_\varepsilon \qquad \lambda_3 > 0 \end{aligned}\right\} \qquad (3.2.3.1)$$

To obtain a reasonable estimate of λ_3 we go through an argument similar to the one used for λ_2. If e_t is very large, of the order of 10σ for example, then it is more likely to be the result of an "outlier" or a large "step change". A "growth change" model is most likely when e_t is of the order of 3σ to 4σ. Suppose therefore that we wish forecast errors larger than 4σ to be given a total probability by $M_t^{(3)}$ which is equal to that given by $M_t^{(1)}$ to forecast errors larger than 2σ, i.e.

$$p(e_t \geqslant 4\sigma \mid M_t^{(3)}) = p(e_t \geqslant 2\sigma \mid M_t^{(1)})$$

which is true when

$$2\sqrt{V^{(3)}} = 4\sqrt{V^{(1)}}$$

or $$\sqrt{V_\varepsilon + \lambda_3 V_\varepsilon} = 2\sqrt{V_\varepsilon}$$

leading to $\lambda_3 = 3$.

3.2.4. Model of a "step change" state : $M_t^{(4)}$

From equations (3.2.2) and (3.2.3) it can be seen that an unusually large $\delta\mu_t$ perturbation will cause a permanent shift in the level μ_t without disturbing the growth β_t. This situation is illustrated at point D of figure 3.1 where a "step change" has taken place but theprocess growth is subsequently unaffected. A "step change" model $M_t^{(4)}$ should therefore view the forecast error e_t as made up of (i) a usual observation noise contribution ε_t, plus (ii) an unusually large $\delta\mu_t$ contribution, with (iii) $\delta\beta_t = 0$ since a non-zero $\delta\beta_t$ would imply a change in the process growth which is not the case.

Hence e_t is viewed by $M_t^{(4)}$ as having been generated by:

$$\left.\begin{array}{llll} & e_t \sim N(0, V^{(4)}) & & \\ \text{with} & V_\varepsilon^{(4)} = V_\varepsilon & & \\ & V_\mu^{(4)} = \lambda_4 V_\varepsilon & \lambda_4 > 0 & \\ & V_\beta^{(4)} = 0 & & \end{array}\right\} \qquad (3.2.4.1)$$

Supposing that one wishes the total probability given by $M_t^{(4)}$ to values of $e_t \geq 5\sigma$ to be equal to that given by $M_t^{(1)}$ to values of $e_t \geq 1\sigma$, then

$$p(e_t \geq 5\sigma \mid M_t^{(4)}) = p(e_t \geq 1\sigma \mid M_t^{(1)})$$

which is true when

$$\sqrt{V^{(4)}} \quad = \quad 5\sqrt{V^{(1)}}$$

or
$$\sqrt{V_\varepsilon + \lambda_4 V_\varepsilon} \quad = \quad 5\sqrt{V_\varepsilon}$$

leading to $\lambda_4 = 24$.

3.3. The updating procedure

In the previous section it was seen how $M_t^{(j)}$ defines a state j and characterises the process evolution at time t, by:

$$y_t = \mu_t + \varepsilon_t^{(j)} \qquad \varepsilon_t^{(j)} \sim N(0, V_\varepsilon^{(j)}) \qquad (3.3.1)$$

$$\mu_t = \mu_{t-1} + \beta_t + \delta\mu_t^{(j)} \qquad \delta\mu_t^{(j)} \sim N(0, V_\mu^{(j)}) \qquad (3.3.2)$$

$$\beta_t = \beta_{t-1} + \delta\beta_t^{(j)} \qquad \delta\beta_t^{(j)} \sim N(0, V_\beta^{(j)}) \qquad (3.3.3)$$

for $j = 1,2,3,4$

Hence the DLM representation of the MSM is similar to that given by (2.4.3) for the LGM except that now there are four models for the noise variance V_t and the disturbance matrix $\underline{W}_t$:

$$\left.\begin{aligned}
&n = 2 \quad \underline{F}_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \quad \underline{G} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \\
&(V_t \mid M_t^{(j)}) = V_\varepsilon^{(j)} \\
&(\underline{W}_t \mid M_t^{(j)}) = \begin{bmatrix} V_\mu^{(j)} + V_\beta^{(j)} & V_\beta^{(j)} \\ V_\beta^{(j)} & V_\beta^{(j)} \end{bmatrix}
\end{aligned}\right\}$$

Notation: The superscripts i and j are from now on to be associated with time periods t-1 and t respectively.

Before y_t is known each model $M_{t-1}^{(i)}$ summarises the prior information about the parameter vector and process state in terms of (i) and (ii) below:

(i) $$(\underline{\Theta}_{t-1} \mid D_{t-1}, M_{t-1}^{(i)}) \sim N(\underline{m}_{t-1}^{(i)}, \underline{C}_{t-1}^{(i)}) \qquad (3.3.4)$$

for $i = 1,2,3,4$

where $$\underline{\Theta}_{t-1} = \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \end{bmatrix} \qquad \underline{m}_{t-1}^{(i)} = \begin{bmatrix} m_{t-1}^{(i)} \\ b_{t-1}^{(i)} \end{bmatrix}$$

$$\underline{C}_{t-1}^{(i)} = \begin{bmatrix} c_{11,t-1}^{(i)} & c_{12,t-1}^{(i)} \\ c_{21,t-1}^{(i)} & c_{22,t-1}^{(i)} \end{bmatrix}$$

$$\left.\text{(ii)} \quad p_{t-1}^{(i)} = \begin{array}{l}\text{probability that } M_{t-1}^{(i)} \text{ was the} \\ \text{correct model at time } t-1 \text{ for} \\ i = 1,2,3,4.\end{array}\right\} \quad (3.3.5)$$

Given that the process is in state i at time $t-1$, there are clearly sixteen possible transition models $M^{(ij)}$ for the process moving from state i to state j; $i,j = 1,2,3,4$. The MSM assumes that the probability of transition $\Pi^{(j)}$ to state j, is independent of time and process history. The choice and implications of $\Pi^{(j)}$ are considered in Chapter 5 where we examine the relative importance and interactions of system parameters such as $\Pi^{(j)}$ and λ_2, λ_3 and λ_4.

As soon as y_t becomes available, sixteen normal distributions,

$$(\underline{\Theta}_t \mid D_t, M^{(ij)}) \sim N(\underline{m}^{(ij)}, \underline{C}^{(ij)}) \qquad (3.3.6)$$

associated with transitions i to j can be derived by applying the

Kalman filter given by equations (2.2.5), (2.2.6) and (2.2.7) with:

(i) $\underline{F}_t$, $\underline{G}$ as for the LGM

(ii) $$\left.\begin{aligned} \underline{m}_{t-1} &= \underline{m}_{t-1}^{(i)} \\ \underline{C}_{t-1} &= \underline{C}_{t-1}^{(i)} \end{aligned}\right\} \quad \text{given by (3.3.4) for } i = 1,2,3,4$$

(iii) $$\left.\begin{aligned} V_t &= V_\varepsilon^{(j)} \\ \underline{W}_t &= \begin{bmatrix} V_\mu^{(j)} + V_\beta^{(j)} & V_\beta^{(j)} \\ V_\beta^{(j)} & V_\beta^{(j)} \end{bmatrix} \end{aligned}\right\} \quad \text{for } j = 1,2,3,4$$

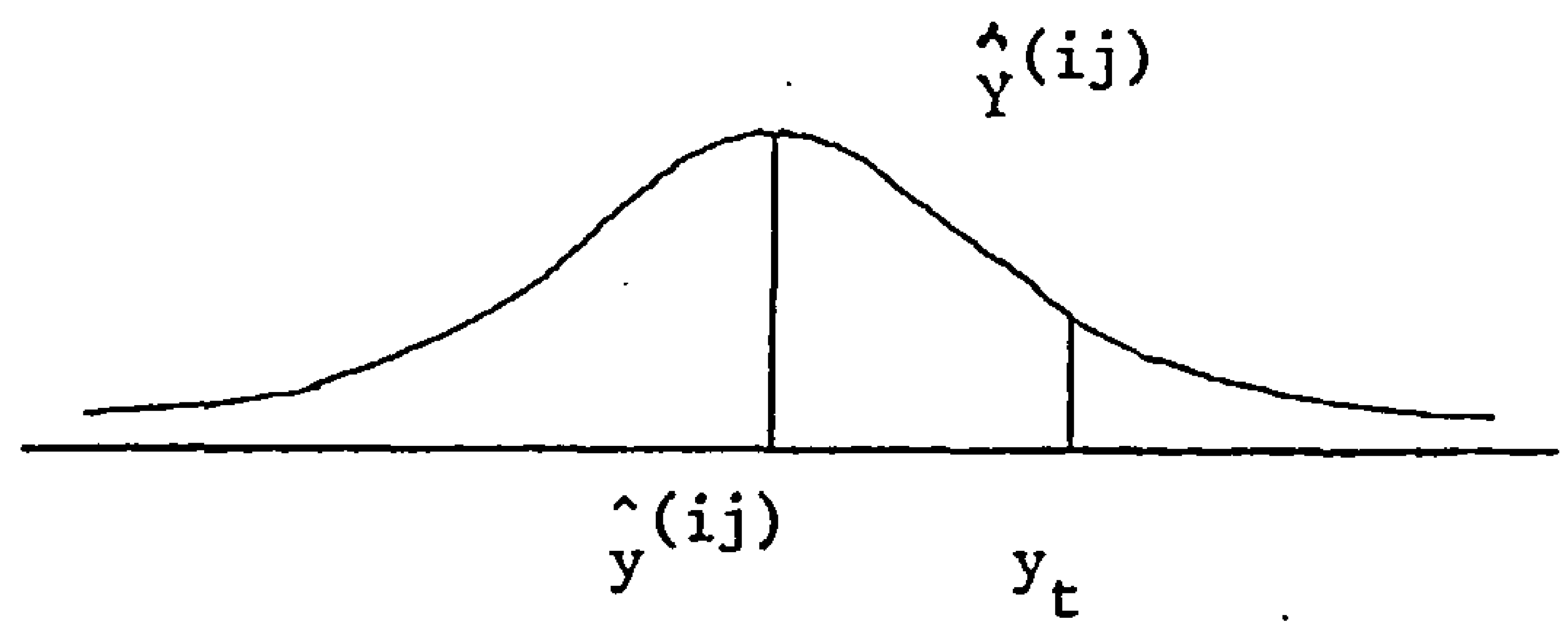
where $\underline{m}^{(ij)} = \begin{bmatrix} m^{(ij)} \\ b^{(ij)} \end{bmatrix}$ and $\underline{C}^{(ij)} = \begin{bmatrix} c_{11}^{(ij)} & c_{12}^{(ij)} \\ c_{21}^{(ij)} & c_{22}^{(ij)} \end{bmatrix}$

are given by:

$$\left.\begin{aligned}
m^{(ij)} &= m_{t-1}^{(i)} + b_{t-1}^{(i)} + A_1^{(ij)} \; e^{(ij)} \\
b^{(ij)} &= b_{t-1}^{(i)} + A_2^{(ij)} \; e^{(ij)} \\
c_{11}^{(ij)} &= r_{11}^{(ij)} \; (1 - A_1^{(ij)}) \\
c_{12}^{(ij)} &= r_{12}^{(ij)} \; (1 - A_2^{(ij)}) \\
C_{22}^{(ij)} &= r_{22}^{(ij)} - \hat{Y}^{(ij)} \; (A_2^{(ij)})^2 \\
\hat{y}^{(ij)} &= m_{t-1}^{(i)} + b_{t-1}^{(i)} \\
e^{(ij)} &= y_t - \hat{y}^{(ij)} \\
r_{11}^{(ij)} &= c_{11,t-1}^{(i)} + 2c_{12,t-1}^{(i)} + c_{22,t-1}^{(i)} + V_\mu^{(j)} + V_\beta^{(j)} \\
r_{12}^{(ij)} &= c_{12,t-1}^{(i)} + c_{22,t-1}^{(i)} + V_\beta^{(j)} \\
r_{22}^{(ij)} &= c_{22,t-1}^{(i)} + V_\beta^{(j)} \\
\hat{Y}^{(ij)} &= r_{11}^{(ij)} + V_\varepsilon^{(j)} \\
A_1^{(ij)} &= r_{11}^{(ij)} / \hat{Y}^{(ij)} \\
A_2^{(ij)} &= r_{12}^{(ij)} / \hat{Y}^{(ij)}
\end{aligned}\right\} \quad (3.3.7)$$

From these updating equations consider now the predictive distribution of y_t as viewed by $M^{(ij)}$:

$$(y_t \mid D_{t-1} , M^{(ij)}) \sim N(\hat{y}^{(ij)} , \hat{Y}^{(ij)})$$



Given the latest observation y_t, the forecast error

$$e^{(ij)} = y_t - \hat{y}^{(ij)}$$

associated with $M^{(ij)}$ can be used together with $\hat{Y}^{(ij)}$ to calculate the likelihood $L^{(ij)}$ of y_t having been generated by the above predictive distribution:

$$L^{(ij)} = L(y_t \mid M^{(ij)}, D_{t-1}) \propto [\hat{Y}^{(ij)}]^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}\frac{[e^{(ij)}]^2}{\hat{Y}^{(ij)}}\right\} \quad (3.3.8)$$

Hence posterior probabilities $p^{(ij)}$ can also be calculated using Bayes' theorem, for $M^{(ij)}$ being the correct transition model in the time interval (t-1 , t):

$$p^{(ij)} = p\,(M^{(ij)} \mid y_t\,, D_{t-1})$$

or $$p^{(ij)} \propto L(y_t \mid M^{(ij)}, D_{t-1}) \cdot p(M^{(ij)} \mid D_{t-1})$$

or $$p^{(ij)} \propto L^{(ij)} \cdot p_{t-1}^{(i)} \cdot \Pi^{(j)} \quad (3.3.9)$$

Passing to time t+1 the updating is still possible by applying the Kalman filter and using $\underline{m}^{(ij)}$, $\underline{C}^{(ij)}$ in place of $\underline{m}_{t-1}^{(i)}$, $\underline{C}_{t-1}^{(i)}$ which then leads to 64 transition models $M^{(ijk)}$. For practical reasons however an approximation procedure is necessary to summarise the parameter and model information in terms of four posterior Normal distributions and four posterior probabilities:

(i) $(\underline{\Theta}_t \mid D_t , M_t^{(j)}) \sim N(\underline{m}_t^{(j)} , \underline{C}_t^{(j)})$ (3.3.10)

(ii) $\left. \begin{array}{l} p_t^{(j)} = p(M_t^{(j)} \mid D_t) = \text{probability that } M_t^{(j)} \\ \quad \text{is the correct model at time } t. \end{array} \right\}$ (3.3.11)

Clearly (3.3.10) and (3.3.11) are then identical to the prior information given by (3.3.4) and (3.3.5) except for the time subscripts and superscripts. Hence they can be used as our time t+1 prior information and the whole updating procedure can be repeated on receipt of y_{t+1} .

To obtain the approximation given by (3.3.10) and (3.3.11) a *collapsing process* has been suggested by H/S [18], which works as follows:

Consider the four transition models to state 1, i.e.

$M^{(i1)}$ for i = 1,2,3,4.

Associated with these four models are probabilities $p^{(i1)}$ (i = 1,2,3,4) and four bivariate normal distributions as given by (3.3.6)

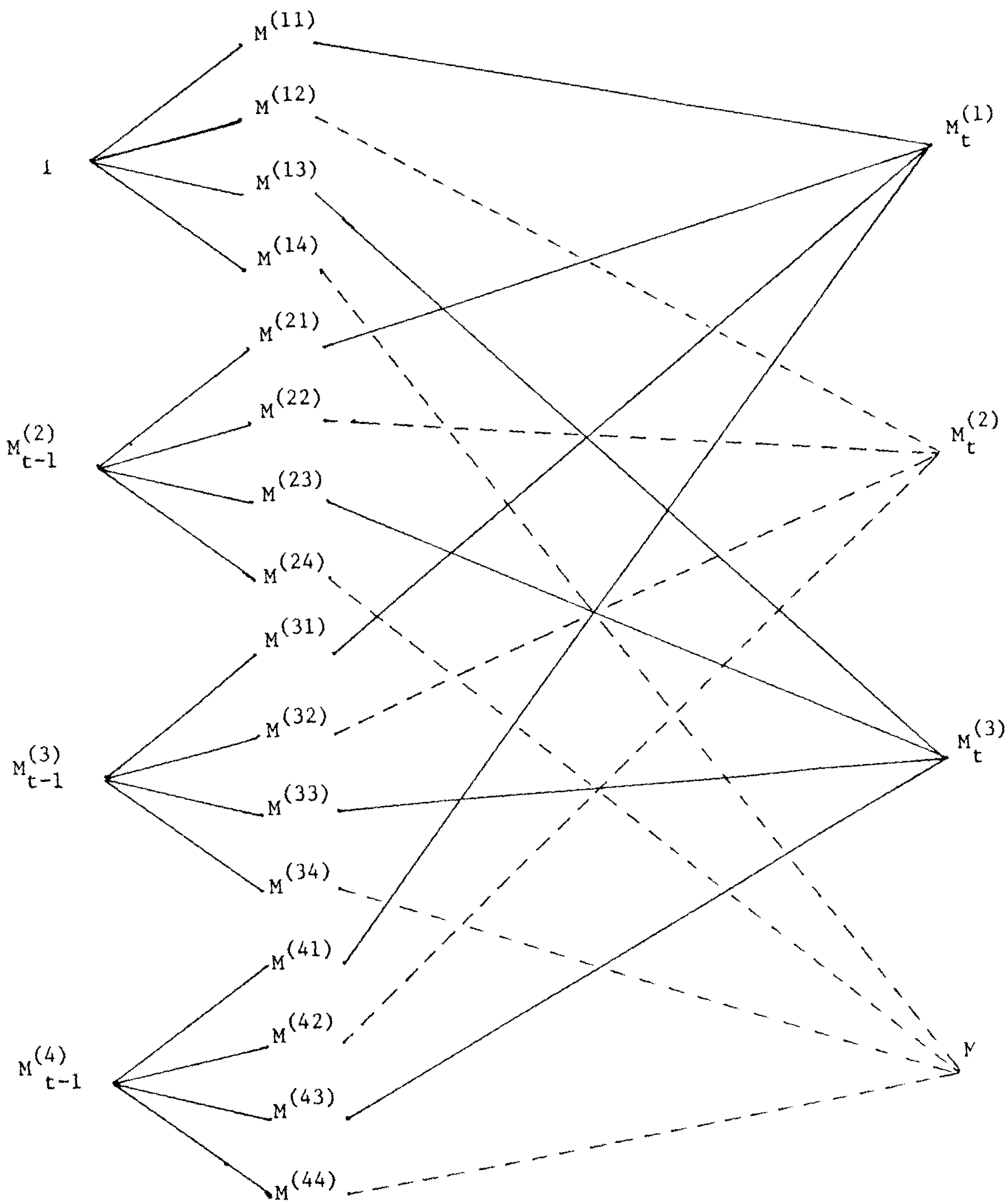
$$(\underline{\theta}_t \mid M^{(i1)}, D_t) \sim N(\underline{m}^{(i1)}, \underline{C}^{(i1)}) \qquad i = 1,2,3,4.$$

Consider now a *mixed* distribution resulting from a probability weighted combination of the four component distributions above (i = 1,2,3,4). Let the first and second moments of the mixed distribution be E_1 and E_2. These can be derived along the lines suggested in Appendix E. The collapsing process assumes that a single distribution with mean equal to E_1 and variance equal to E_2 is a good approximation of the mixed distribution and therefore reduces the four distributions associated with $M^{(i1)}$ to a single one.

Similarly for transitions to states 2, 3 and 4, single distributions replace the sets of four bivariate normal distributions associated with:

$$\left.\begin{matrix} M^{(i2)} \\ M^{(i3)} \\ M^{(i4)} \end{matrix}\right\} \qquad \text{for } i = 1, 2, 3, 4.$$

Figure 3.3 illustrates the way transition models are combined and the complete collapsing process is summarised by the set of equations (3.3.12) which are analogous to those derived in Appendix E.



$M^{(11)}$
$M^{(12)}$
$M^{(13)}$
$M^{(14)}$
$M^{(21)}$
$M^{(22)}$
$M^{(23)}$
$M^{(24)}$
$M^{(31)}$
$M^{(32)}$
$M^{(33)}$
$M^{(34)}$
$M^{(41)}$
$M^{(42)}$
$M^{(43)}$
$M^{(44)}$
$M_{t-1}^{(2)}$
$M_{t-1}^{(3)}$
$M_{t-1}^{(4)}$
$M_t^{(1)}$
$M_t^{(2)}$
$M_t^{(3)}$

Fig. 3.3.

$$\left.\begin{aligned}
p_t^{(j)} &= \sum_{i=1}^{4} p^{(ij)} \\
k_i &= p^{(ij)} \,/\, p_t^{(j)} \\
m_t^{(j)} &= \sum_{i=1}^{4} k_i \; m^{(ij)} \\
b_t^{(j)} &= \sum_{i=1}^{4} k_i \; b^{(ij)} \\
c_{11,t}^{(j)} &= \sum_{i=1}^{4} k_i \; \{c_{11}^{(ij)} + [m^{(ij)} - m_t^{(j)}]^2 \} \\
c_{12,t}^{(j)} &= \sum_{i=1}^{4} k_i \; \{c_{12}^{(ij)} + [m^{(ij)} - m_t^{(j)}]\,[b^{(ij)} - b_t^{(ij)}]\} \\
c_{22,t}^{(j)} &= \sum_{i=1}^{4} k_i \; \{c_{22}^{(ij)} + [b^{(ij)} - b_t^{(j)}]^2\}
\end{aligned}\right\} \qquad (3.3.12)$$

for j = 1, 2, 3, 4

3.4. Supplying the initial information

3.4.1. Introduction

In Section 1.4 it was stressed that an important advantage of Bayesian models is their ability to start forecasting in situations where there is little or no prior data history. In the MSM it is necessary to specify three sets of values expressing our beliefs and expectations about the series in question prior to the first observation at time t = 1.

(i) STARTING VALUES for $\underline{m}_o^{(i)}$ $\underline{C}_o^{(i)}$ and $p_o^{(i)}$; $i = 1, 2, 3, 4$. These are required initially by the Kalman updating procedure which produces revised estimates $\underline{m}_t^{(i)}$, $\underline{C}_t^{(i)}$ and $p_t^{(i)}$ posterior to the latest observation y_t.

(ii) NORMAL NOISE VARIANCE which is an estimate of the true but unknown process observation noise variance V_ε. Since this is an initially nominated estimate of V_ε, it will be denoted by $V_{\varepsilon,N}$ consistent with the notation used earlier for the SSM and LGM of chapter 2.

(iii) SYSTEM PARAMETERS which are used to determine the model variances $V_\varepsilon^{(j)}$, $V_\mu^{(j)}$, $V_\beta^{(j)}$ and the transition probabilities $\Pi^{(j)}$.

Subsections 3.4.2., 3.4.3. and 3.4.4. will now consider (i), (ii) and (iii) in greater detail.

3.4.2. Starting values for $\underline{m}_o^{(i)}$ $\underline{C}_o^{(i)}$ and $p_o^{(i)}$

These represent our beliefs about the parameter vector $\underline{\theta}_o$ as viewed by the four models,

$$(\underline{\theta}_o \mid M_o^{(i)}) \sim N(\underline{m}_o^{(i)} \quad \underline{C}_o^{(i)}) \qquad i = 1, 2, 3, 4.$$

as well as about the likely state of the process,

$$p_o^{(i)} = p(M_o^{(i)}) \qquad i = 1, 2, 3, 4.$$

These may be specified as follows for all i :

$$\underline{m}_o^{(i)} = \begin{bmatrix} m_o \\ b_o \end{bmatrix} \quad ; \quad \underline{C}_o^{(i)} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \quad ; \quad p_o^{(i)} = \Pi^{(i)}$$

where

m_o = the best estimate for the level μ_o

b_o = the best estimate for the growth β_o

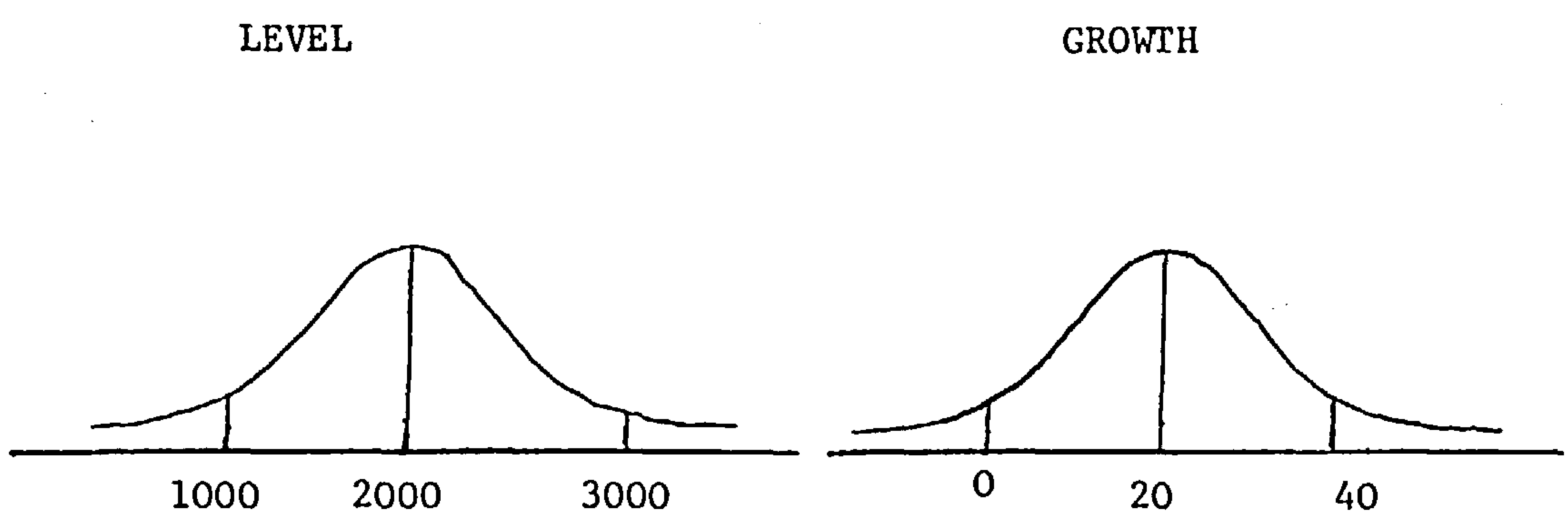
c_{11} = variance on the estimate of the level

c_{22} = variance on the estimate of the growth

$c_{12} = c_{21}$ = covariance of the level and growth estimates

$\Pi^{(i)}$ = probability of process transition to state i

The approach to be taken in order to arrive at these estimates is now outlined. Consider a product launch situation where management believes that the level of sales for their product is likely to be between 1000 and 3000 units a month with an expected growth between zero and 2% per month but have no insight as to the correlation between level and growth. This subjective information can reasonably be expressed in terms of the following probability distributions:



where 0, 20, 40 are 0%, 1% and 2% of the expected level (2000), respectively. If we now assume that the ranges shown cover approximately four standard deviations then the following starting values can be specified:

$$m_o = 2000$$

$$b_o = 20$$

$$c_{11} = 2500$$

$$c_{12} = c_{21} = 0$$

$$c_{22} = 100$$

Finally $p_o^{(i)}$ for $i = 1, 2, 3, 4$ can be set equal to $\Pi^{(i)}$ so that the initial set of prior probabilities for the state of the process corresponds to the expected long run relative frequencies of the process states. Note however that $p_t^{(i)}$, $\underline{m}_t^{(i)}$ and $\underline{C}_t^{(i)}$ are updated at each point in time as described by the updating procedure of section 3.2. and therefore the choice of the starting values m_o, b_o, c_{11}, c_{12}, c_{22} and $p_o^{(i)}$ is not critical since it affects the behaviour of the MSM only during the very early stages.

3.4.3. Normal noise variance

This represents our belief about the variability of the process in its "no change" state. If for example in passing from time $t - 1$ to t the process is in a "no change" state then the implication is that the underlying true level and growth parameters have not changed considerably so that,

$$\mu_t \sim \mu_{t-1} \text{ and } \beta_t \sim \beta_{t-1}$$

and therefore the forecast error is simply the result of the observation noise ε_t reflecting the customer ordering variations which are assumed to be random, such that

$$\varepsilon_t \sim N\ (0, V_\varepsilon)$$

In order to arrive at an estimate $V_{\varepsilon,N}$ of V_{ε}, management has to answer the following question: what is the maximum acceptable forecast error e_{max} which can be explained by pure random noise without giving rise to suspicions about a possible "change" situation. If this maximum "residual" is assumed to be roughly two standard deviations in size then an estimate of V_{ε} can be obtained from:

$$e_{max} = 2\sqrt{V_{\varepsilon,N}}$$

The robustness of the MSM to errors in the assumed variability ($V_{\varepsilon,N}$) will be examined in chapter 5 and methods for its on line estimation are proposed in chapter 6.

3.4.4. System parameters

In section 3.2 it was shown how the MSM can model different states using distinct combinations of $V_{\varepsilon}^{(j)}$, $V_{\mu}^{(j)}$ and $V_{\beta}^{(j)}$ in a number of models $M_t^{(j)}$. These variances given by equations (3.2.1.1), (3.2.2.1), (3.2.3.1) and (3.2.4.1) can be summarised in table 3.1 below. Note that $V_{\varepsilon,N}$ is used instead of V_{ε} since the latter is an unknown true parameter and therefore in practice the different states have to be modelled using its initially nominated estimate $V_{\varepsilon,N}$.

	STATE 1 "no change" j = 1	STATE 2 "outlier" j = 2	STATE 3 "growth change" j = 3	STATE 4 "step change" j = 4
$V_{\varepsilon}^{(j)}$	$V_{\varepsilon,N}$	$\lambda_2 V_{\varepsilon,N}$	$V_{\varepsilon,N}$	$V_{\varepsilon,N}$
$V_{\mu}^{(j)}$	0	0	0	$\lambda_4 V_{\varepsilon,N}$
$V_{\beta}^{(j)}$	0	0	$\lambda_3 V_{\varepsilon,N}$	0

TABLE 3.1

Associated with model $M_t^{(j)}$ is a transition probability $\Pi^{(j)}$ (j = 1, 2, 3, 4) which was used in section 3.3 for the updating of the information in the MSM. Hence apart from $V_{\varepsilon,N}$ there are seven parameters $\{\lambda_2, \lambda_3, \lambda_4, \Pi^{(1)}, \Pi^{(2)}, \Pi^{(3)}, \Pi^{(4)}\}$ which are necessary in order to model the likely environment which is going to affect th data series in question, where by "environment" we mean both the different process states and their relative frequency of occurrence. This set of parameters will therefore be referred to from now on as *system parameters* and the abbreviation SSP will be used to stand for Set of System Parameters.

$$\text{i.e.} \quad SSP = \{\lambda_2, \lambda_3, \lambda_4, \Pi^{(1)}, \Pi^{(2)}, \Pi^{(3)}, \Pi^{(4)}\} \qquad (3.4.4.1)$$

H/S [19] have suggested a set of values for the system parameters which using the present notation is equivalent to the following SSP;

λ_2	λ_3	λ_4	$\Pi^{(1)}$	$\Pi^{(2)}$	$\Pi^{(3)}$	$\Pi^{(4)}$
101	1	100	0.9	0.094	0.003	0.003

TABLE 3.2.

The discrepancy between the H/S values for $\{\lambda_2, \lambda_3, \lambda_4\}$ = {101, 1, 100} and the set {100, 3, 24} derived earlier in section 3.2. , does not seem to affect the performance of the system critically as will be suggested by the results of chapter 5.

The choice of SSP controls the performance of the system especially at points of discontinuities and in chapter 5 it will be shown how the response to outliers and the speed of response to abrupt growth and step changes is affected by varying different parameters in the SSP. Before that however, it is necessary to consider the problem of measuring the performance of the MSM.

CHAPTER 4

Measures of Performance

The aim of this chapter is to examine the problems of assessing the results produced by the MSM and comparing the behaviour of the system produced as a result of different choices of SSP. Traditional comparative criteria of optimality such as the MSE are no longer adequate and additional measures of performance must be adopted. This presents a problem of great difficulty and one acceptable although lengthy procedure including graphical presentations is proposed. This will subsequently be used as the basic language of understanding and judging the responses of the system.

4.1 The problem

The type of process modelled by a MSM may contain outliers as well as sudden changes in the underlying true level and growth. Under these circumstances if the MSE of the one step ahead system forecasts is used as a criterion of performance it will be misleading since errors resulting from the occasional large disturbances at points of discontinuities will be the major contributors to its final value.

This can be illustrated in terms of a simple example. Consider an artificially generated process with observation noise variance $V_\varepsilon = 0.0025$ (the generating model is described in Section 4.2) containing an outlier of size 10σ at time $t = 21$, a 1σ change in growth at

$t = 41$ and a 10σ step change at $t = 61$. The performance of the MSM for two choices of SSP on this series, is shown in Figure 4.1. The two SSP's and the MSE over the whole period are shown in Table 4.1.

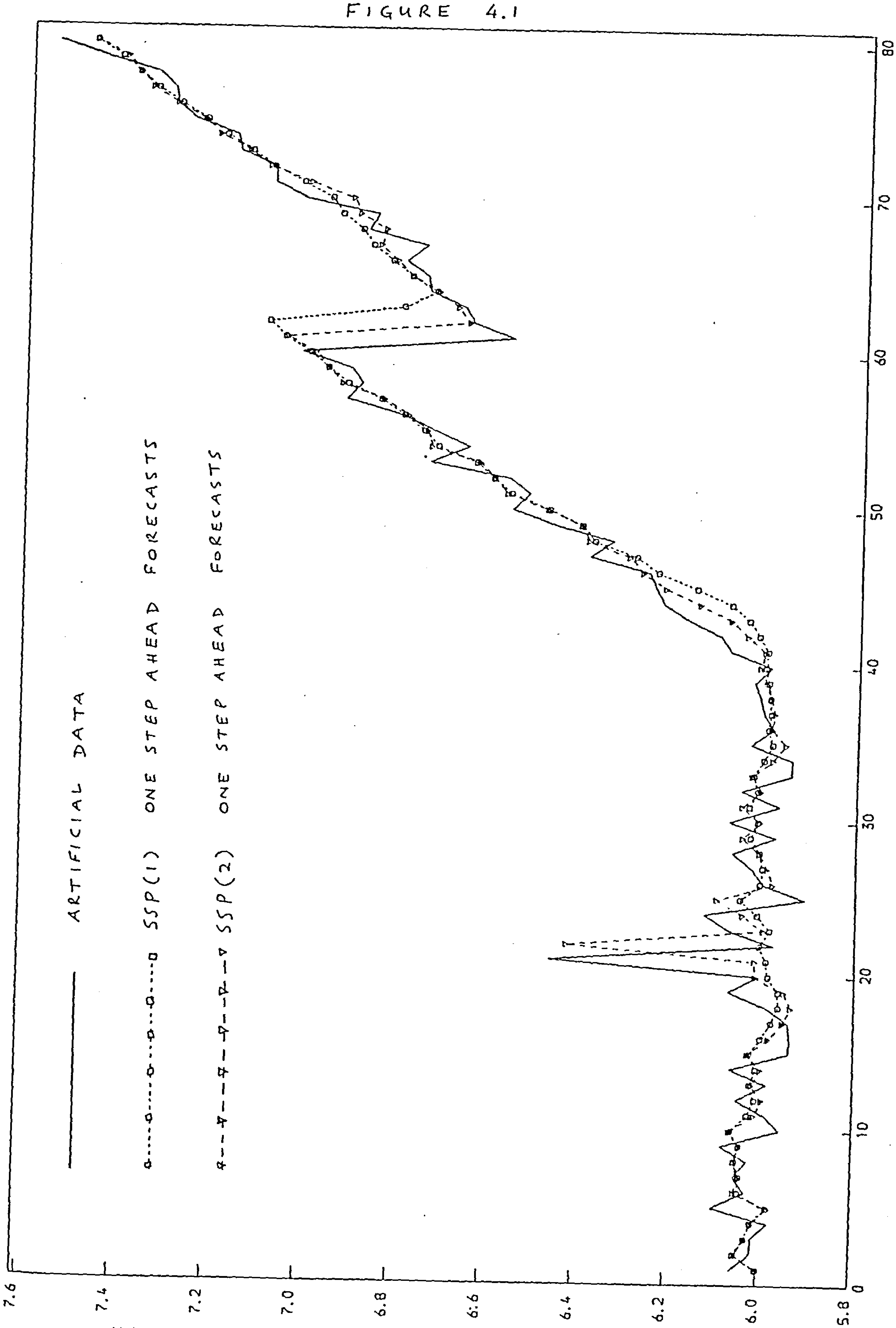
	λ_2	λ_3	λ_4	$\Pi^{(1)}$	$\Pi^{(2)}$	$\Pi^{(3)}$	$\Pi^{(4)}$	MSE
SSP(1)	101	1	100	.9	.094	.003	.003	.0120
SSP(2)	101	1	100	.9	.004	.048	.048	.0114

TABLE 4.1

Note that SSP(1) is the SSP equivalent to the set of system parameters suggested by H/S [19] based on what they call "basic precepts" and taking into consideration a set of utilities associated with the different types of forecasting errors. This will from now on be referred to as the *standard* SSP and will be used as a starting point for the sensitivity analysis of Chapter 5.

From Figure 4.1 it can be seen that the performance of the standard SSP is quite different from that of SSP(2) especially at points of discontinuities. SSP(1) treats an unusually large forecast error as an outlier unless a step change is confirmed by subsequent observations. In contrast SSP(2) is very responsive, therefore more unstable during quiet periods and responds in an undesirable way to outliers. Yet if the MSE criterion is used in the traditional sense then SSP(2) is selected (see Table 4.1) mainly because of the "wait and see" SSP(1) response to step changes. In most stock control and other

FIGURE 4.1


ARTIFICIAL DATA
SSP(1) ONE STEP AHEAD FORECASTS
SSP(2) ONE STEP AHEAD FORECASTS
7.6
7.4
7.2
7.0
6.8
6.6
6.4
6.2
6.0
5.8
0
10
20
30
40
50
60
70
80

applications however the "wait and see" approach is desirable since spot orders occur frequently and they should be treated as outliers and effectively ignored. The MSE comparison is in fact not only misleading but also meaningless since it does not allow any qualitative comparisons of the form "this SSP choice responds too slowly to growth changes" or "responds badly to outliers" etc.

The point is that no single criterion can possibly reflect the differences in response after points of discontinuities such as growth and step changes. Yet the true test of performance for any system is its responses immediately after points of "change" and therefore the performance of the MSM should be viewed as being four dimensional with each dimension somewhow measuring the system's response corresponding to one of the four modelled process states.

Recalling that STATES 1, 2, 3 and 4 correspond to "no change", "outlier", "growth change" and "step change" events, it is convenient to denote the response corresponding to a particular state j by $R^{(j)}$ for $j = 1,2,3,4$. The way we propose to measure $R^{(j)}$ in order to arrive at a final definition of performance will now be considered in the next section.

4.2. Measuring the goodness of $R^{(j)}$

First it is necessary to define three terms which will often be used from now on:

(i) *system forecast* $\hat{y}_{t+1}$

(ii) *system level* m_t

(iii) *system growth* b_t

It was seen in section 3.3 of the previous chapter that the updating procedure summarises the posterior information at time t in terms of,

$$\left.\begin{aligned} p_t^{(j)} &= p(M_t^{(j)} \mid D_t) \\ \left(\begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} \middle| D_t , M_t^{(j)} \right) &\sim N\left(\begin{bmatrix} m_t^{(j)} \\ b_t^{(j)} \end{bmatrix} ; \underset{\sim}{C}_t^{(j)} \right) \end{aligned}\right\} \text{ for } j = 1,2,3,4$$

By *system level* and *system growth* we mean the expectation of μ_t and β_t conditional only on D_t and therefore independent of $M_t^{(j)}$. These are given by the probability weighted averages of $m_t^{(j)}$ and $b_t^{(j)}$:

$$\text{System level } m_t = E(\mu_t \mid D_t) = \sum_j p_t^{(j)} \; m_t^{(j)}$$

$$\text{System growth } b_t = E(\beta_t \mid D_t) = \sum_j p_t^{(j)} \; b_t^{(j)}$$

and based on these the one step ahead *system forecast* is given by:

$$\text{System forecast} \quad \hat{y}_{t+1} = m_t + b_t$$

In order to successfully implement a Bayesian MSM it is necessary to understand the response of the system to different choices of SSP in relation to different types and sizes of discontinuities. With the usual "messy" data of the real world however it is impossible to analyse and assess these responses because the characteristics of the underlying true process such as the observation noise variance V_ε and the level and growth parameters μ_t, β_t are unknown. As a result our estimates of these true parameters can not be easily judged. It is therefore necessary to generate artificial data series where all the characteristics are known thus allowing comparisons of m_t, b_t with the true process parameters μ_t and β_t. The generating models for the different process states j and the way we propose to measure the system's response $R^{(j)}$ to each state is now described.

4.2.1. Response to state 1 - "no change" : $R^{(1)}$

$$\left.\begin{aligned} \text{Let } \mu_o &= 6 \\ \beta_o &= 0 \\ r_t &\sim N(0,1) \quad \text{generated using a NAG subroutine (see Appendix F)} \\ V_\varepsilon &= \sigma^2 = 0.0025 \end{aligned}\right\} \qquad (4.2.1.1)$$

Then the model used for generating a period of "no change" can be written as follows:

$$\left.\begin{aligned} y_t &= \mu_t + r_t\sigma \\ \mu_t &= \mu_{t-1} + \beta_t \\ \beta_t &= \beta_{t-1} \end{aligned}\right\} \quad t \geqslant 1$$

One such realisation of 200 points is shown in Figure 4.2. The y_t process generated has 2σ range from 5.9 to 6.1 and for a multiplicative model representing real log-normally distributed data Y_t, it implies a 2σ range for Y_t from exp(5.9) to exp(6.1), that is from approximately 365 to 446 units of the "product" in question.

Because of the stable structure implied by state 1, $R^{(1)}$ can adequately be measured by the traditional MSE criterion, and in this context the MSE can be viewed as simply one dimension of performance, measuring the goodness of the system in the absence of discontinuities. Hence given a choice of SSP, $R^{(1)}$ will be described by:

$$\text{MSE} = (1/200) \sum_{t=1}^{200} (y_t - \hat{y}_t)^2$$

where $\hat{y}_t$ is the system forecast for time period t made at time t-1. The lower the MSE the better $R^{(1)}$ will be considered to be since it implies smaller forecast errors. For the two SSP choices considered earlier in Table 4.1, the MSE as defined here is 0.00249 and 0.00302 respectively. Hence although from Figure 4.1 SSP(2) can be

FIGURE 4.2

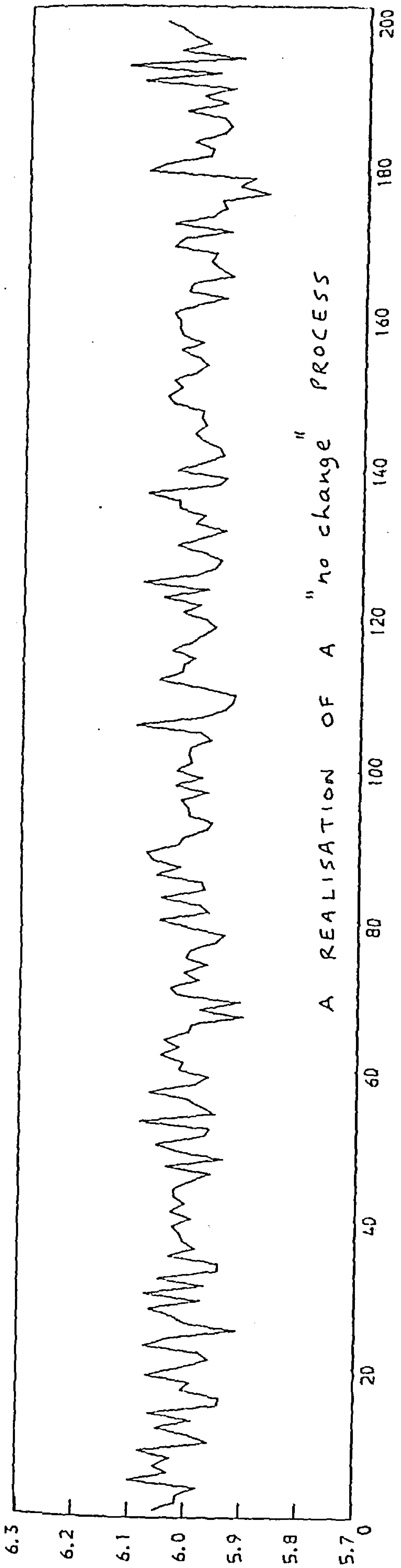
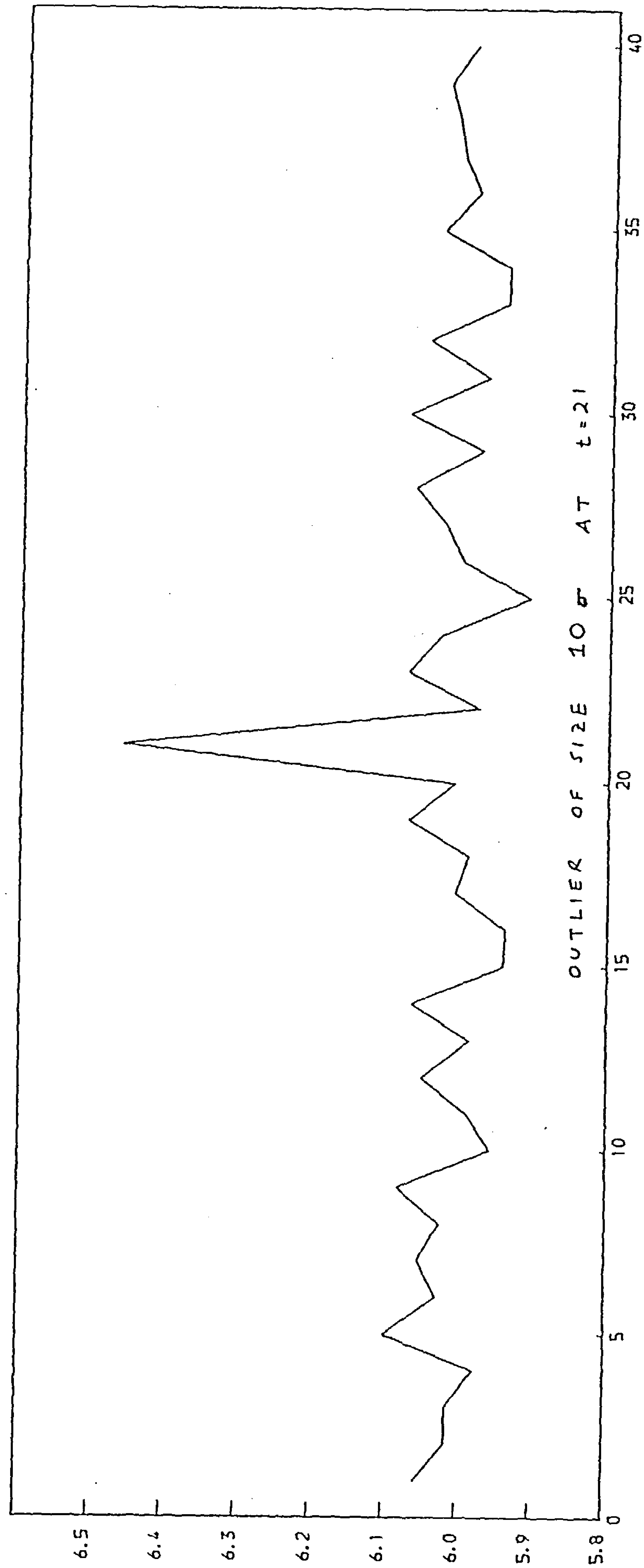


FIGURE 4.3



seen to respond faster to growth changes, its MSE is approximately 21% higher for periods of "no change".

4.2.2. Response to state 2 - "outlier" : $R^{(2)}$

The generating model is the same except for the observation equation: .

$$\left.\begin{aligned} y_t &= \mu_t + r_t\sigma + \delta_t k_2 \sigma \\ \mu_t &= \mu_{t-1} + \beta_t \\ \beta_t &= \beta_{t-1} \end{aligned}\right\} \quad \begin{aligned} & t \geqslant 1 \\ & \delta_t = \begin{cases} 0 & t \neq t^* \\ 1 & t = t^* \end{cases} \end{aligned}$$

This introduces a freak observation of size $k_2\sigma$ at time $t = t^*$. One realisation generated from this model with $t^* = 21$ and $k_2 = 10$ or in other words with a 10σ outlier at time $t = 21$, is shown in Figure 4.3.

The response $R^{(2)}$ to an outlier generated in this way at time t, can easily be described by examining the difference $\hat{y}_{t+1} - \hat{y}_t$ and expressing it in terms of standard deviations σ.

$$\text{Let } \hat{y}_{t+1} - \hat{y}_t = z\sigma$$

Then if z is sufficiently small the response of the system is considered to be satisfactory since the smaller the z value the smaller the effect of the outlier on the future course of the system.

For the outliers of Figure 4.1 at $t = 21$ the z values corresponding to SSP(1) and SSP(2) are given below:

	$\hat{y}_{21}$	$\hat{y}_{22}$	$z = (\hat{y}_{22} - \hat{y}_{21}) \ / \ 0.05$
SSP(1)	5.992	6.007	0.3
SSP(2)	6.017	6.423	8.12

TABLE 4.2

Response to the outlier of Figure 4.1.

Clearly the z values describe $R^{(2)}$ very well, with $z = 0.3$ indicating that SSP(1) effectively ignores the 10σ outlier while SSP(2) responds badly by feeding approximately 80% of the 10σ forecast error (resulting from the outlier at $t = 21$) to the next period forecase $\hat{y}_{22}$.

A crude decision rule is that $z > 1$ (i.e. difference $\hat{y}_{t+1} - \hat{y}_t$ greater than 1σ) implies an undesirable response $R^{(2)}$ placing too much weight on a freak observation which ideally should be ignored.

4.2.3. Response to state 3 - "growth change" : $R^{(3)}$

The following generating model is used to introduce a growth change of size $k_3\sigma$ at time $t = t^*$

$$\left.\begin{aligned} y_t &= \mu_t + r_t\sigma \\ \mu_t &= \mu_{t-1} + \beta_t \\ \beta_t &= \beta_{t-1} + \delta_t k_3\sigma \end{aligned}\right\} \quad \begin{aligned} & t \geq 1 \\ & \\ & \delta_t = \begin{cases} 0 & t \neq t^* \\ 1 & t = t^* \end{cases} \end{aligned}$$

One realisation generated from this model with $t^* = 21$ and $k_3 = 1$ is shown in Figure 4.4 together with the one step ahead system forecast $\hat{y}_t$ corresponding to the two SSP's of Table 4.1.

We are interested in the response of the system immediately after the growth change since after approximately ten time periods the system adapts itself and its behaviour is then well described by the "no change" response, $R^{(1)}$. Unlike $R^{(1)}$ and $R^{(2)}$ which are well described by the MSE and z respectively, $R^{(3)}$ can not adequately be described by a quantitative criterion since a single number can not reflect the variety of possible responses over the period following an abrupt growth change. We therefore propose to illustrate it graphically in terms of the system growth b_t which is the relevant system response at a growth change and can be compared with the true β_t parameter.

For the example of Figure 4.4, β_t is zero until $t = 20$ jumping to a new level of $1\sigma = 0.05$ from $t = 21$ onwards. The b_t response to this discontinuity in β_t is shown in Figure 4.5 over the period of interest.

FIGURE 4.4

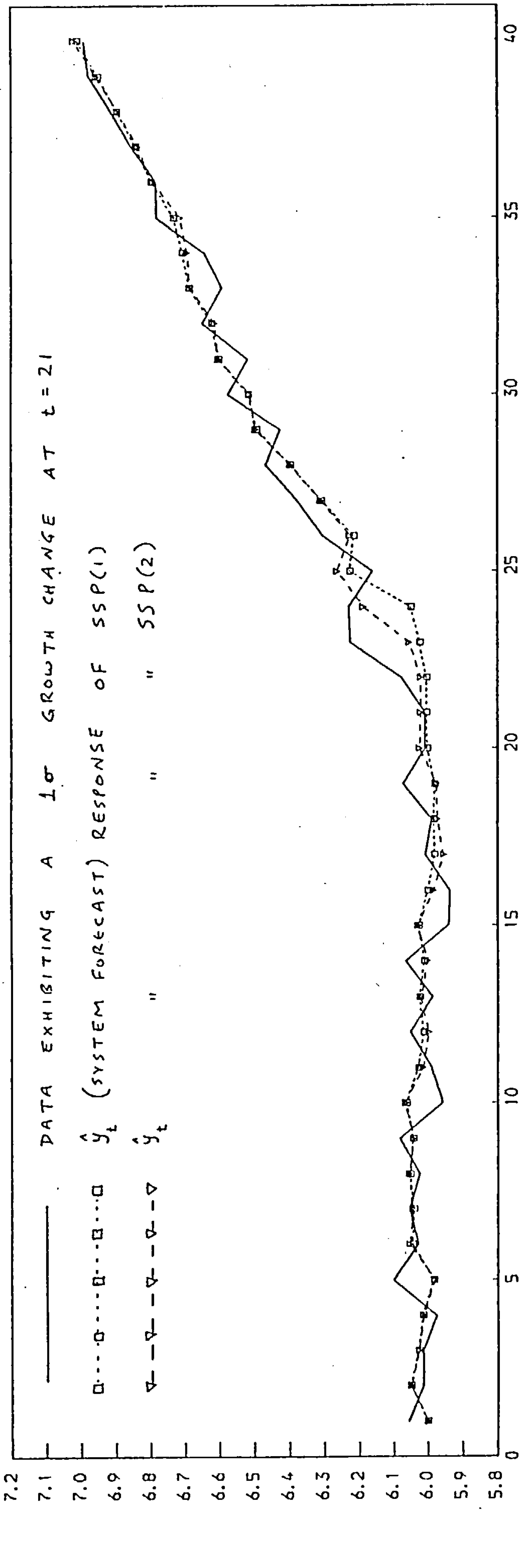
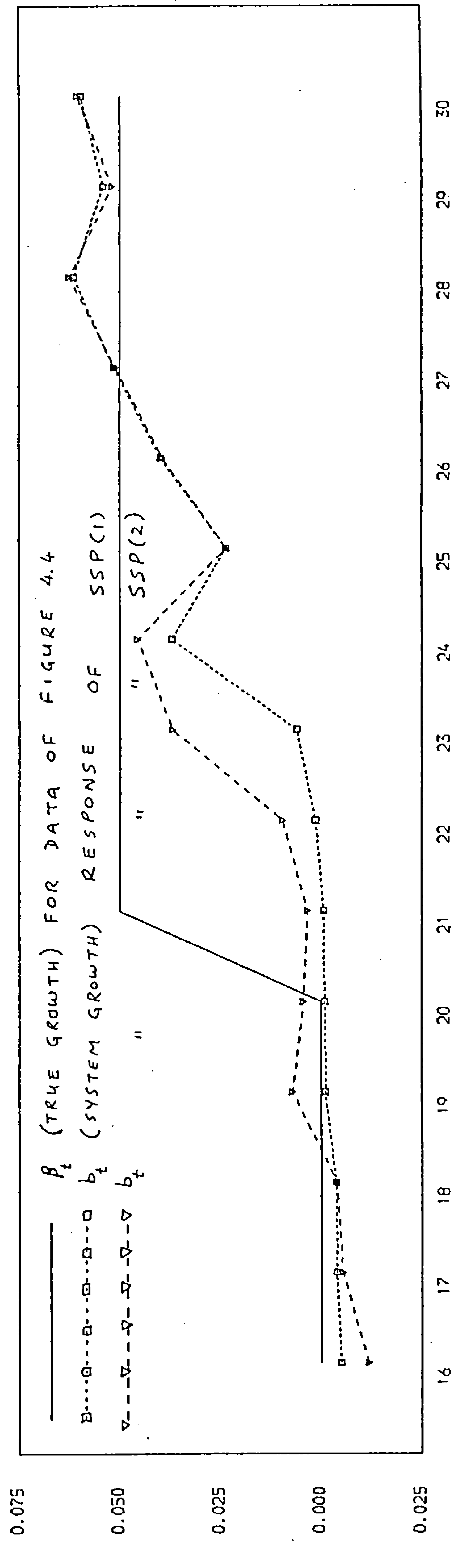


FIGURE 4.5



4.2.4. Response to state 4 - "step change" : $R^{(4)}$

Finally the generating model used to introduce a "step change" of size $k_4\sigma$ at time $t = t^*$ is as follows:

$$\left.\begin{aligned} y_t &= \mu_t + r_t\sigma \\ \mu_t &= \mu_{t-1} + \beta_t + \delta_t k_4\sigma \\ \beta_t &= \beta_{t-1} \end{aligned}\right\} \quad \begin{aligned} & t \geq 0 \\ & \\ & \delta_t = \begin{cases} 0 & t \neq t^* \\ 1 & t = t^* \end{cases} \end{aligned}$$

One realisation exhibiting a step change of size 5σ at time $t = 21$ ($k_4 = -5$, $t^* = 21$) is shown in Figure 4.6.

Like $R^{(3)}$ the only satisfactory way of assessing $R^{(4)}$ is qualitatively using a graph over the period following the discontinuity of the true parameter μ_t. For a change in μ_t however the relevant response is the system level m_t which estimates μ_t and can be compared with it as shown in Figure 4.7 for the 5σ step change of Figure 4.6.

4.3. Consistency of performance

4.3.1. Foreword

The general problem is to differentiate and evaluate SSP's. However measures of system performance can be expected to be very dependent on the actual realisation of the data which could then lead to incorrect conclusions about the expected or average responses of a

FIGURE 4.6

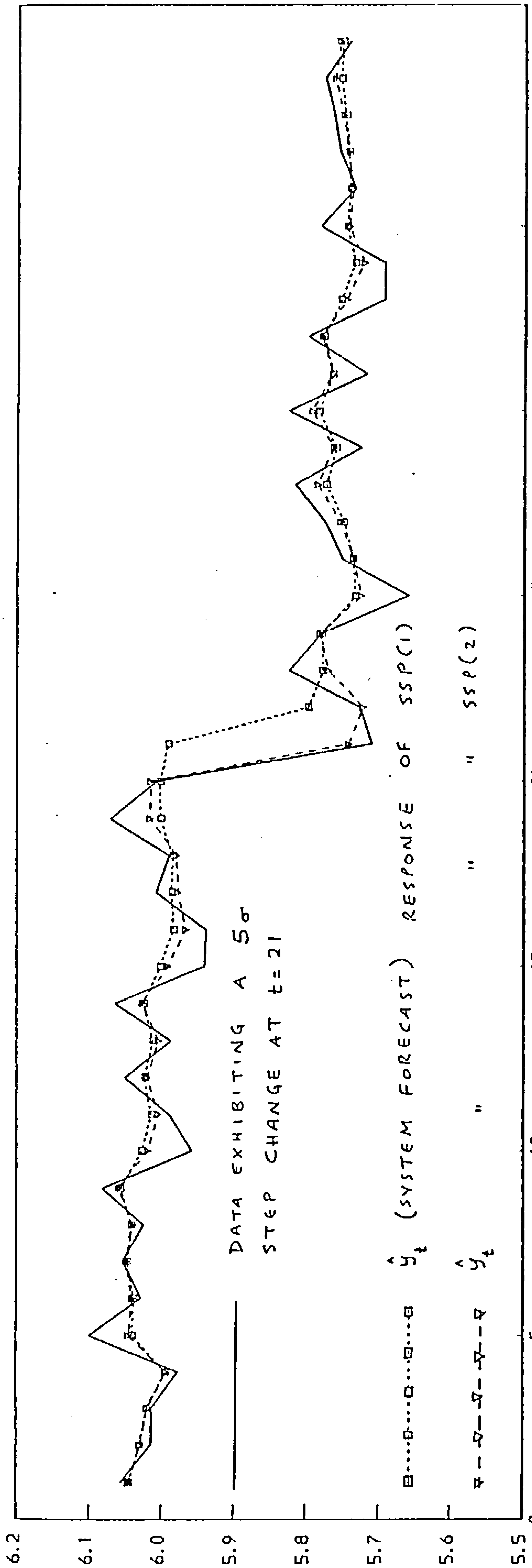
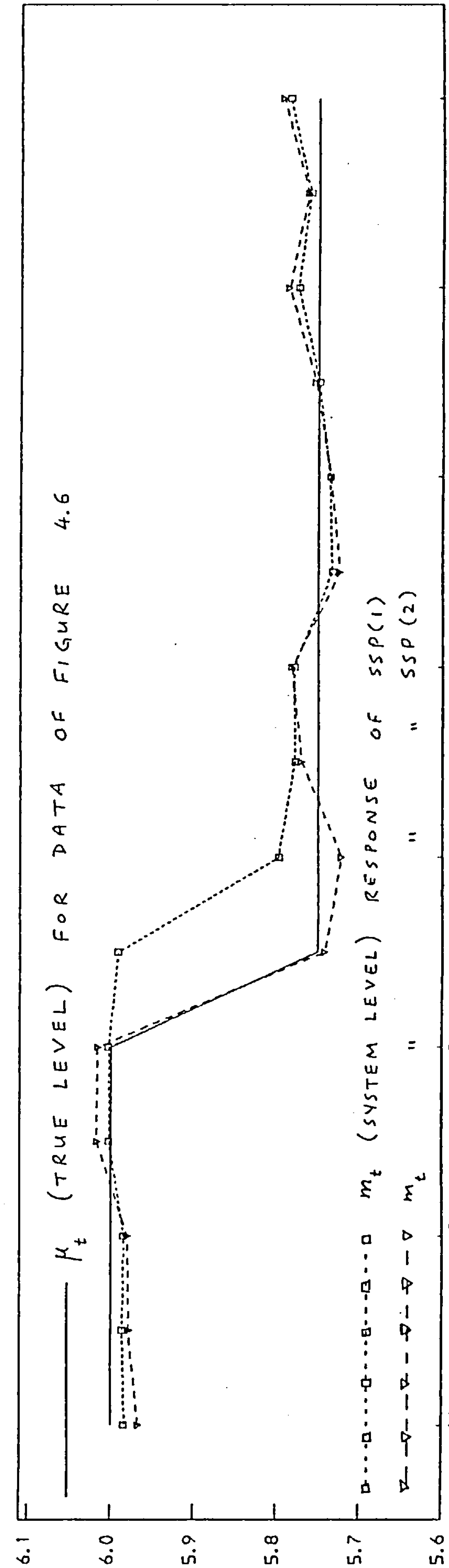


FIGURE 4.7



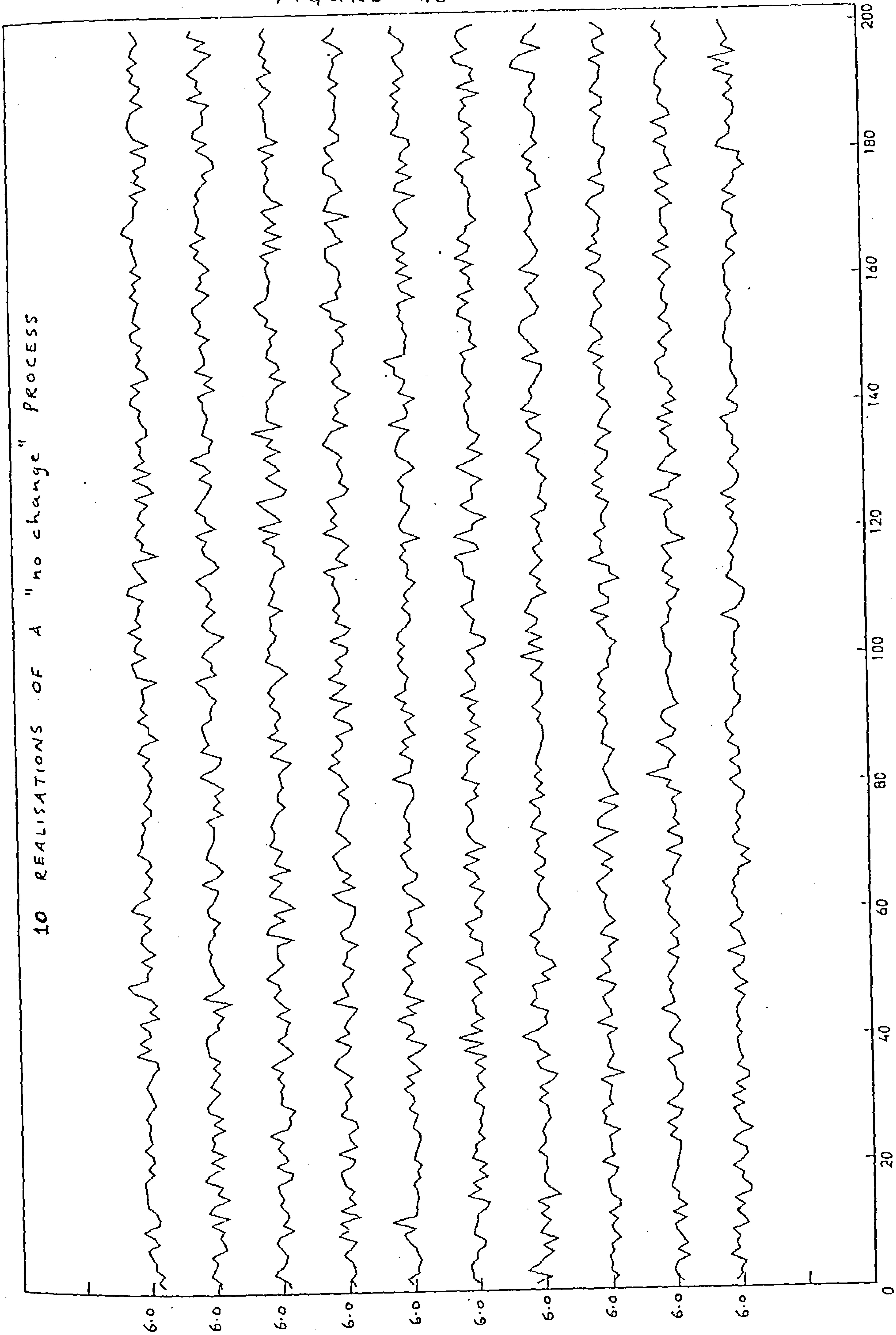
particular SSP. This section demonstrates three things.

First, that different realisations of the data do produce different performance given the same choise of SSP. Second, that whilst there is a difference between realisations, the relative performance between SSP's appears to be consistent across realisations, i.e. if on one realisation a particular SSP responds to a step change faster than another SSP then the first SSP seems to respond faster to a step change on the other realisations. Third, we attempt to illustrate that at points of discontinuity the expected or average responses of an SSP to a lot of realisations is very similar to the responses of the SSP when presented with a single purely deterministic process exhibiting the equivalent discontinuities. This final point has important implications since it is then possible to differentiate and evaluate the "change" responses of different SSP's ($R^{(2)}$, $R^{(3)}$, $R^{(4)}$) on a single purely deterministic process, instead of calculating the average responses across a large number of realisations. Finally it must be stressed that although the above observations are difficult to prove theoretically, their validity is clearly suggested by our simulation results.

4.3.2. Experiment

Consider the ten realisations $\{y_t\}$ $t = 1, 2, \ldots 200$ shown in Figure 4.8 and generated by the "no change" model of 4.2.1 with an observation noise variance $V_\varepsilon = 0.0025$. Consider also the following

FIGURE 4.8


10 REALISATIONS OF A "no change" PROCESS
0
20
40
60
80
100
120
140
160
180
200
6.0
6.0
6.0
6.0
6.0
6.0
6.0
6.0
6.0
6.0

SSP's:

	λ_2	λ_3	λ_4	$\Pi^{(1)}$	$\Pi^{(2)}$	$\Pi^{(3)}$	$\Pi^{(4)}$
SSP(1)	101	1	100	.9	.094	.003	.003
SSP(2)	101	1	100	.7	.025	.250	.025
SSP(3)	101	1/4	100	.9	.094	.003	.003
SSP(4)	101	4	100	.9	.094	.003	.003

In section 3.4.4. a table was given summarising the way that a particular SSP in conjunction with $V_{\varepsilon,N}$ (an initially nominated estimate of V_ε which is unknown in real data) is used to model the environment of the process under observation. From now on we will denote the pair $\left\{SSP,\ V_{\varepsilon,N}\right\}$ by ICS which stands for Initial Condition Set since both $V_{\varepsilon,N}$ and the seven parameters of an SSP must be specified initially. In the present experiment we consider the following 5 ICS's:

ICS(1) : SSP(1) $V_{\varepsilon,N} = .0025$

ICS(2) : SSP(2) $V_{\varepsilon,N} = .0025$

ICS(3) : SSP(1) $V_{\varepsilon,N} = .0100$

ICS(4) : SSP(3) $V_{\varepsilon,N} = .0025$

ICS(5) : SSP(4) $V_{\varepsilon,N} = .0025$

In other words V_ε is assumed known in all cases except for ICS(3) which overestimates V_ε by a factor of four. The choice of these five ICS's has been made because they produce a wide range of responses to the different discontinuities.

The performance $R^{(j)}$ for $j = 1,2,3,4$ of these ICS's over the ten realisations is shown in Tables 4.3, 4.4 and Figures G.1, G.2 (see Appendix G), respectively. Table 4.3 shows the MSE($R^{(1)}$) response as defined earlier in 4.2.1 while Table 4.4 shows the z response ($R^{(2)}$) for a 10σ outlier at $t = 21$.

	ICS(1)	ICS(2)	ICS(3)	ICS(4)	ICS(5)
Realisation 1	.00236	.00359	.00226	.00231	.00234
" 2	.00269	.00428	.00259	.00263	.00268
" 3	.00264	.00436	.00255	.00258	.00263
" 4	.00319	.00472	.00302	.00313	.00319
" 5	.00320	.00530	.00309	.00312	.00320
" 6	.00292	.00463	.00285	.00285	.00291
" 7	.00329	.00515	.00312	.00318	.00329
" 8	.00334	.00533	.00319	.00325	.00335
" 9	.00312	.00489	.00299	.00304	.00311
" 10	.00277	.00421	.00268	.00273	.00277

TABLE 4.3

MSE response – $R^{(1)}$

	ICS(1)	ICS(2)	ICS(3)	ICS(4)	ICS(5)
Realisation 1	.293	4.477	.383	.282	.293
" 2	.353	4.820	.410	.349	.353
" 3	.357	4.781	.491	.346	.369
" 4	.310	5.037	.335	.300	.309
" 5	.325	5.923	.328	.322	.325
" 6	.282	5.444	.296	.292	.280
" 7	.328	4.441	.444	.326	.329
" 8	.314	4.781	.444	.305	.320
" 9	.384	5.434	.381	.375	.385
" 10	.378	4.558	.512	.377	.381

TABLE 4.4

z response – $R^{(2)}$

In Figure G.1 (Appendix G) we show the b_t response $(R^{(3)})$ to a 1σ growth change at time $t = 21$ where,

$$\left.\begin{array}{lll} \beta_t = 0 & t < 21 \\ \beta_t = .05 & t \geq 21 \end{array}\right\} \quad \text{since } \sigma = \sqrt{V_\varepsilon} = .05$$

Figure G.1 consists of ten pages, one for each realisation. Three graphs are included in each page showing the b_t response of three pairs of ICS's as follows:

Top Graph
- ———— β_t the true underlying growth parameter
- .... b_t response of ICS(1)
- - - - - - - b_t " " ICS(2)

Middle Graph
- ———— β_t
- .... b_t response of ICS(1)
- - - - - - - b_t " " ICS(3)

Bottom Graph
- ———— β_t
- .... b_t response of ICS(4)
- - - - - - - b_t " " ICS(5)

Finally the m_t response $(R^{(4)})$ for a 5σ step change where,

$$\left.\begin{array}{ll} \mu_t = 6.00 & t < 21 \\ \mu_t = 5.75 & t \geq 21 \end{array}\right\}$$

is shown in Figure G.2 (see Appendix G) which otherwise has the same "format" as Figure G.1.

4.3.3. Discussion of Results

A careful examination of the results leads to the following two conclusions.

(i) Given a particular realisation then the relative response of any two choices of ICS has been found to be consistent in almost all the realisations considered.

From Table 4.3 for example the MSE response of the five ICS's can be ranked in ascending order of MSE value as follows:

RANK	ICS
1	3
2	4
3	5
4	1
5	2

This ranking is consistent from one realisation to the next and hence given any pair of ICS's the one with a higher MSE in one realisation will have a higher MSE on any other realisation.

From Table 4.4, $R^{(2)}$ is also consistent since no matter which realisation we examine the conclusion we will draw is that the only ICS which responds badly (large z) to outliers is ICS(2).

A third and final example is from Figure G.1 (see Appendix G) and considering the bottom graph in each page. The relative response is clearly consistent from one realisation to the next with ICS(5) always picking up the growth discontinuity mush faster than ICS(4).

(ii) Given a particular choice of ICS then the responses associated with it are to a large extent realisation dependent.

From Table 4.3 for example and for ICS(1), the MSE ranges from .00236 given realisation 1 to .00334 given realisation 8 the latter MSE value being approximately 40% higher. Alternatively looking again at the bottom graph of Figure G.1 it can be seen that ICS(5) reaches the new β_t level in four time periods given realisation 1 and in eight time periods given realisation 10.

The implication of these two conclusions is that although we could consideronly asingle random realisation and analyse the relative responses of the system, this could give us misleading information about the expected or average speed of response to a discontinuity in β_t or μ_t. The obvious approach to this problem is to repeat every exp eriment on many realisations and calculate the average b_t and m_t responses which would then be used for comparisons. Instead it can be shown that a much easier and computationally more efficient alternative is to examine the b_t and m_t responses of the system to a purely deterministic set up as the following experiment illustrates.

4.3.4. Experiment

Consider 100 realisations of a stochastic process y_t, $t=1,2...30$ exhibiting a growth change of size 1σ at time $t = 21$. The generating model for such a process was given earlier in 4.2.3.

Let $\overline{b}_t(1)$ be the average b_t response of ICS(1) to these 100 process realisations, i.e.

$$\overline{b}_t(1) \quad = \quad \Sigma b_t(1) \quad \text{for} \quad t=1,2...30$$

the sum being taken over all realisations.

Similarly let $\overline{m}_t(1)$ be the average m_t response of ICS(1) to the same set of 100 process realisations this time exhibiting a step change of size 5σ at time $t = 21$.

Having defined $\overline{b}_t(1)$ and $\overline{m}_t(1)$ we will now define two more responses, $b_t^*(1)$ and $m_t^*(1)$ associated with the deterministic set up. Consider the following purely deterministic y_t process for $t=1,2...30$ with a 1σ growth change at $t = 21$ as before:

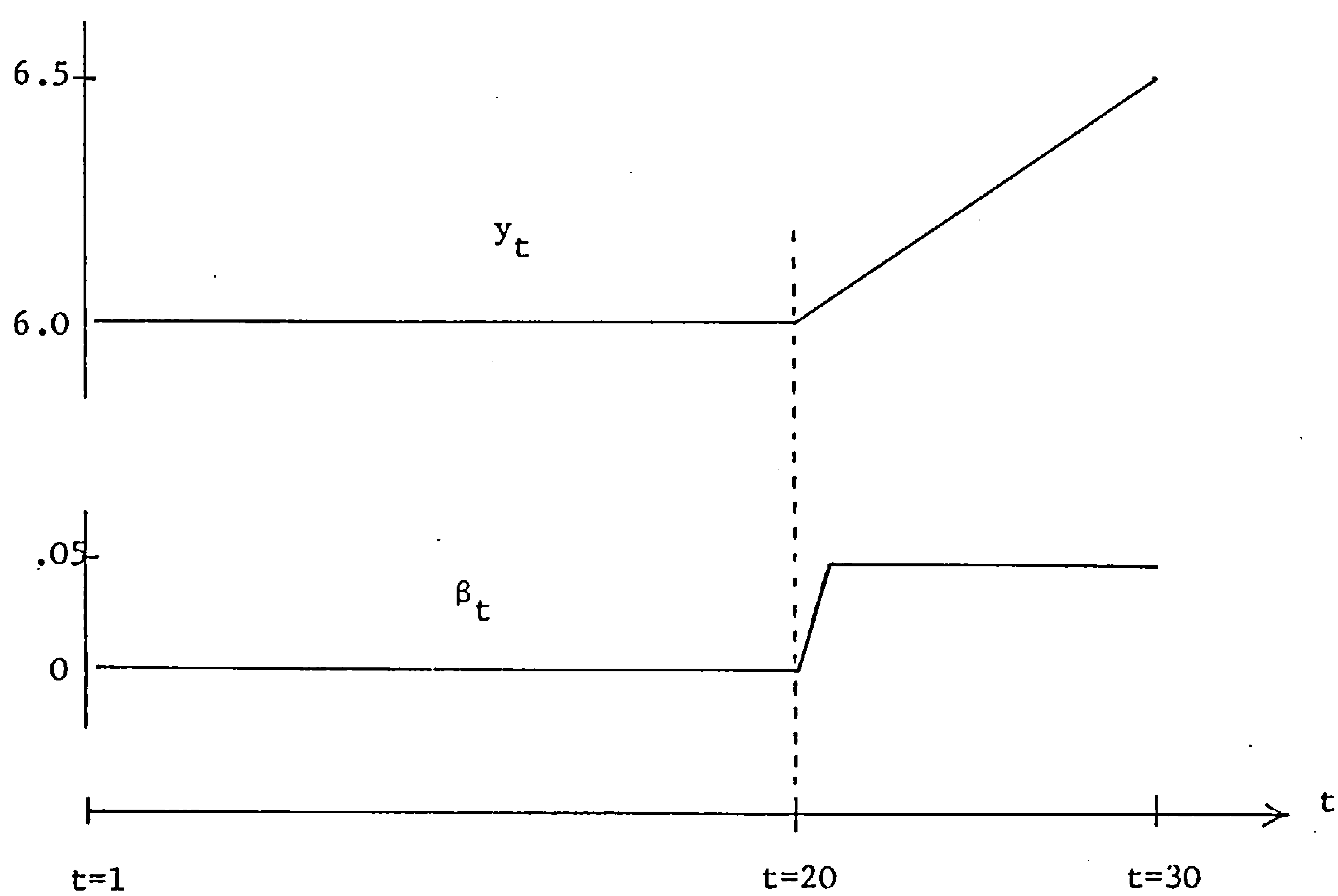
$$\left.\begin{aligned} y_t &= \mu_t \\ \mu_t &= \mu_{t-1} + \beta_t \\ \beta_t &= \beta_{t-1} + \delta_t k_3 \sigma \end{aligned}\right\}$$

with $k_3 = 1$ and $\delta_t = \begin{cases} 0 & t \neq 21 \\ 1 & t = 21 \end{cases}$

$\mu_0 = 6$

$\beta_0 = 0$

This generates the following deterministic process:



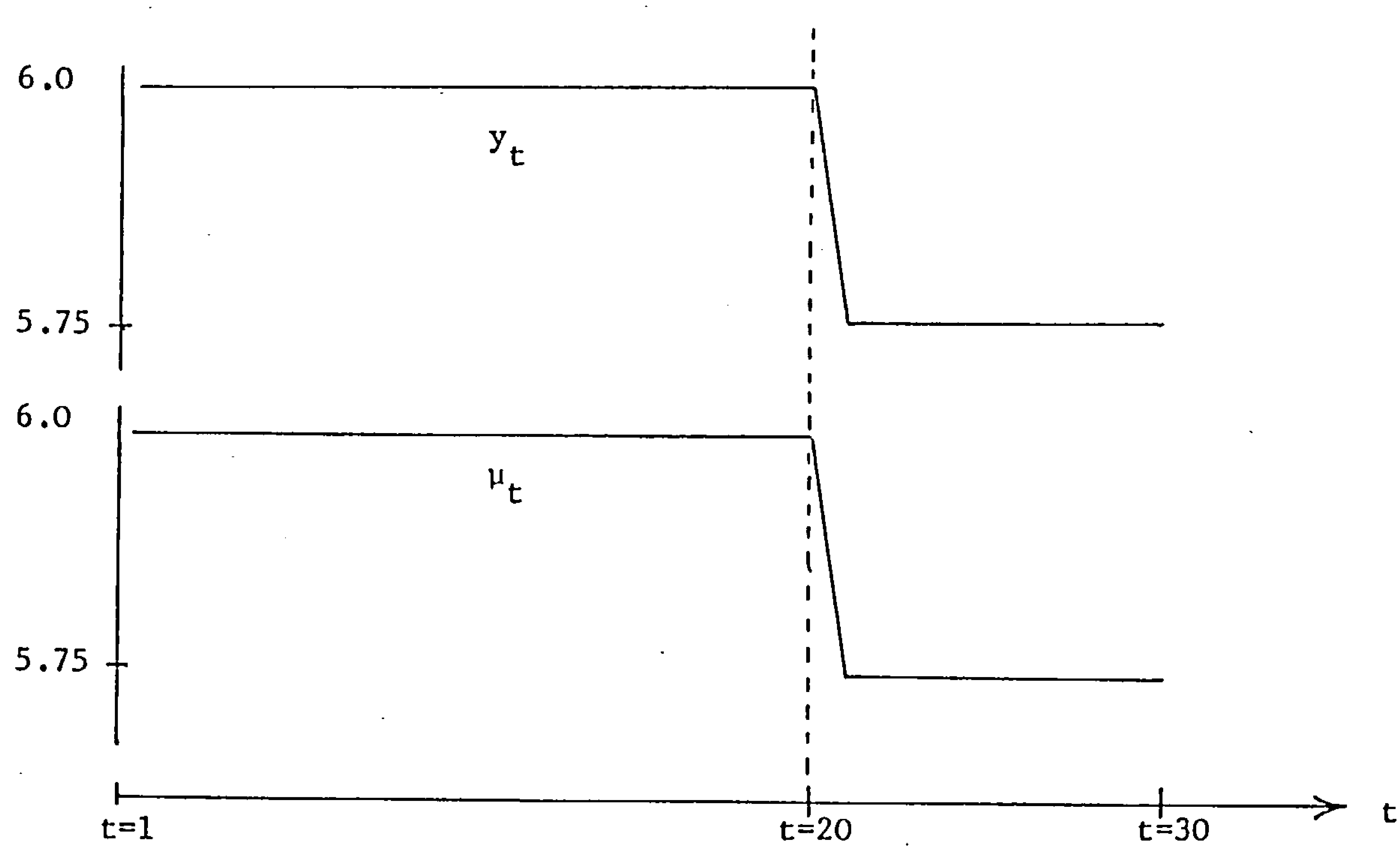
Let $b_t^*(1)$ denote the b_t response of the system given ICS(1) to this deterministic process y_t.

Similarly let $m_t^*(1)$ denote the m_t response of the system given ICS(1) to another deterministic process y_t with a 5σ step change at $t = 21$ and generated by:

$$\left.\begin{aligned} y_t &= \mu_t \\ \mu_t &= \mu_{t-1} + \beta_t + \delta_t \, k_4 \, \sigma \\ \beta_t &= \beta_{t-1} \end{aligned}\right\}$$

$$\text{with} \quad \begin{aligned} \mu_0 &= 6 \\ \beta_0 &= 0 \\ k_4 &= -5 \end{aligned} \qquad \text{and} \quad \delta_t = \begin{cases} 0 & t \neq 21 \\ 1 & t = 21 \end{cases}$$

This generates the following deterministic process observation y_t and process level μ_t:



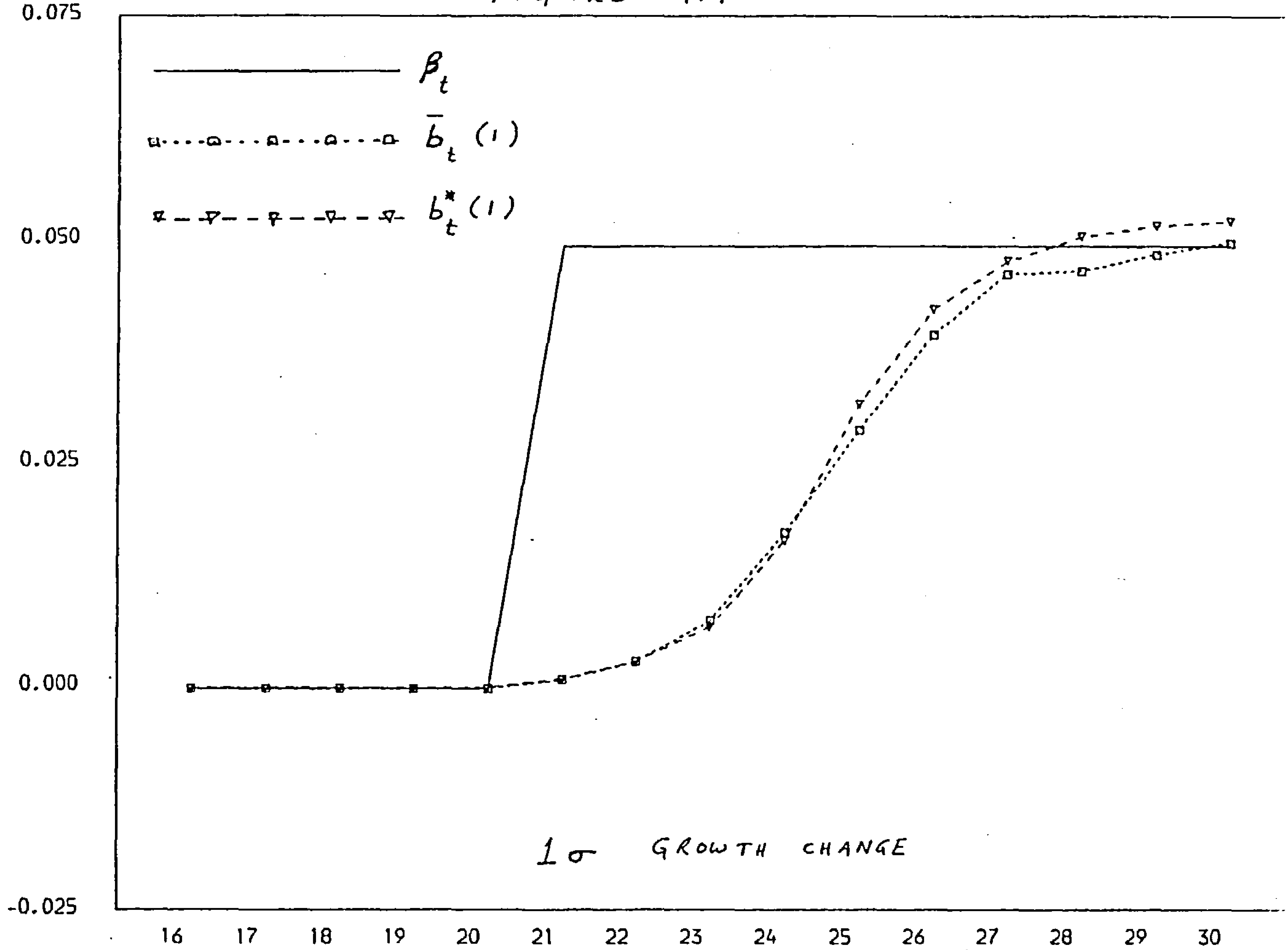
From Figure 4.9 it can be seen that $b_t^*(1)$ and $m_t^*(1)$ are practically identical to $\bar{b}_t(1)$ and $\bar{m}_t(1)$ respectively and when the same experiment is repeated given ICS(3) (see Figure 4.10) and all other ICS's, the results lead to the same conclusion.

Our results therefore suggest that instead of looking at a large number of realisations and studying the average responses $\bar{b}_t$ and $\bar{m}_t$ it is much easier and practically equivalent to examine b_t^* and m_t^* over a single deterministic process exhibiting identical growth and step changes.

The b_t^* response of the five ICS's considered in the previous experiment of 4.3.2, is shown in Figure 4.11 which has the same "format" as Figure G.1 but is much more powerful because it can be viewed as a "summary" of the 10 graphs of Figure G.1 each showing b_t corresponding to a particular realisation. Figure 4.11 however shows b_t without any bias due to realisation since b_t^* is realisation independent and equivalent to the average b_t over an infinite number of realisations.

Similarly Figure 4.12 shows m_t^* of the five ICS's and has the same "format" as Figure G.2 (Appendix G) but the m_t^* response is now realisation independent.

FIGURE 4.9


0.075
0.050
0.025
0.000
-0.025
β_t
$\bar{b}_t$ (1)
b^*_t (1)
1σ GROWTH CHANGE
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

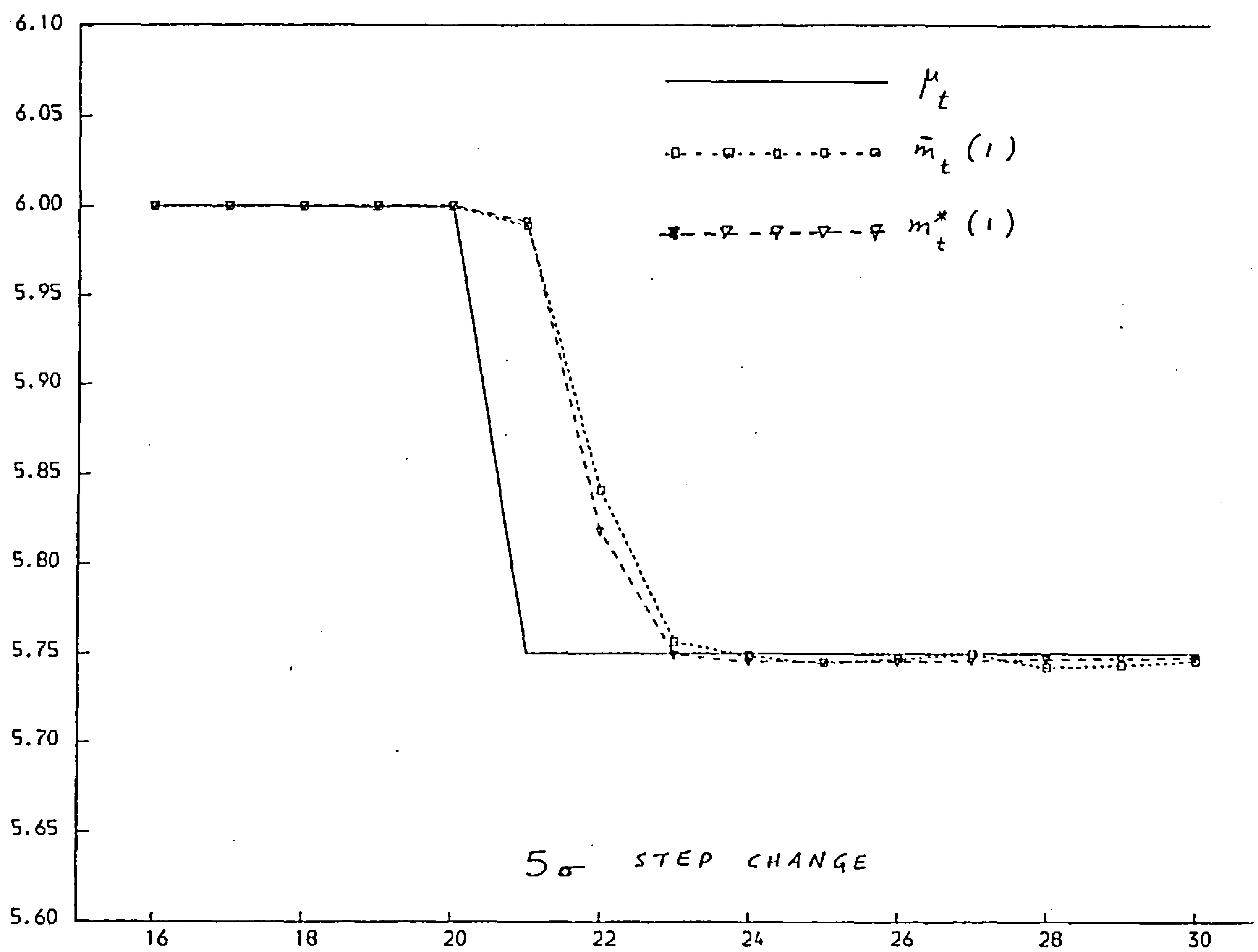
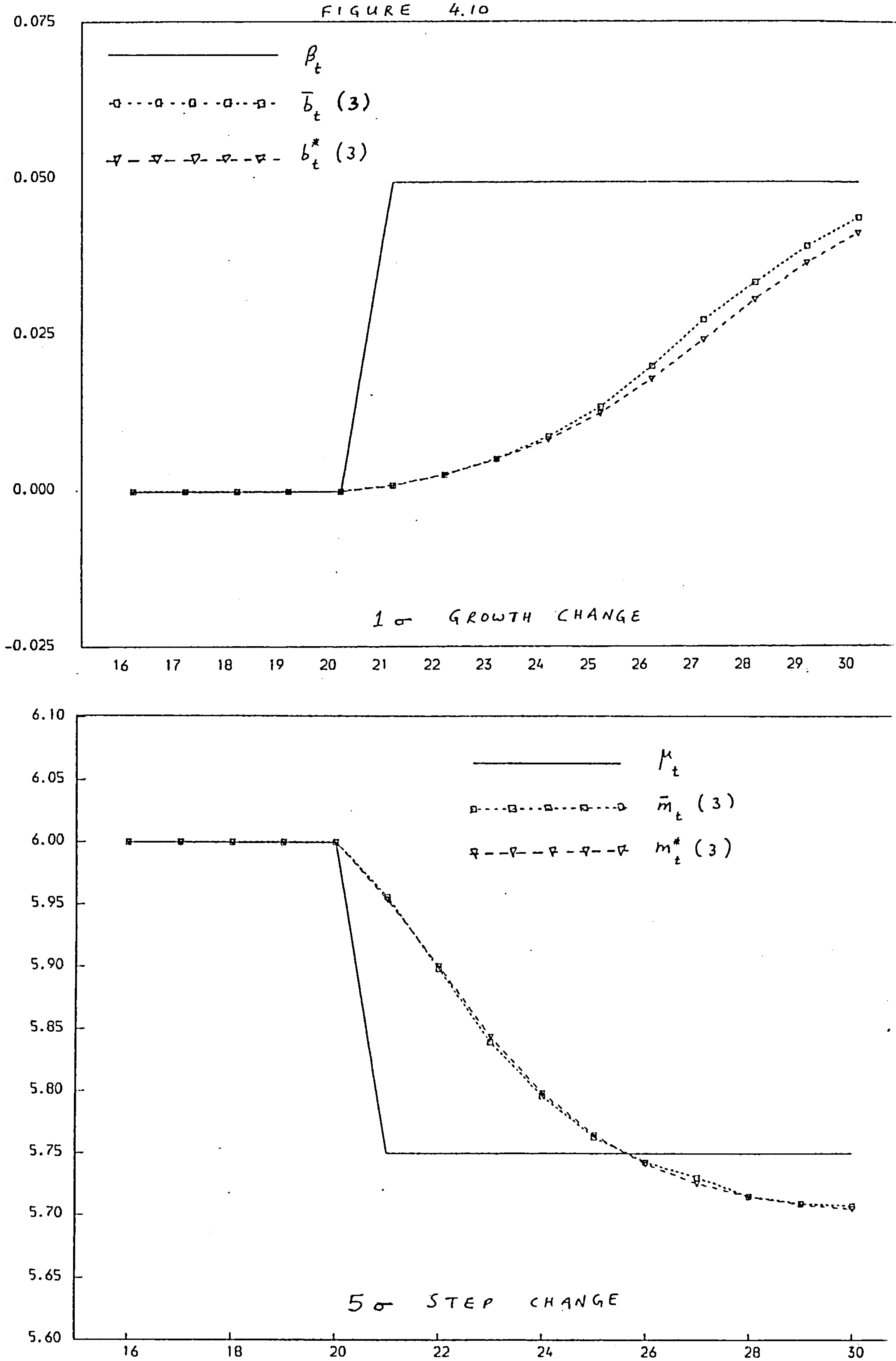

6.10
6.05
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
5.60
μ_t
$\bar{m}_t$ (1)
m^*_t (1)
5σ STEP CHANGE
16
18
20
22
24
26
28
30

FIGURE 4.10



β_t
$\bar{b}_t$ (3)
b_t^* (3)
0.075
0.050
0.025
0.000
-0.025
16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
1 σ GROWTH CHANGE
μ_t
$\bar{m}_t$ (3)
m_t^* (3)
6.10
6.05
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
5.60
16 18 20 22 24 26 28 30
5 σ STEP CHANGE

FIGURE 4.11

b_t^* RESPONSE ASSOCIATED WITH THE ICS's OF SECTION 4.3.2

FIGURE 4.12

m_t^* RESPONSE ASSOCIATED WITH THE ICS's OF SECTION 4.3.2

4.4. Final definition of performance

It is now possible to arrive at the following definition of performance:

$$Performance \;=\; f(R^{(1)}, R^{(2)}, R^{(3)}, R^{(4)})$$

where

(i) $R^{(1)}$ will be measured in terms of the MSE criterion given a stochastic process realisation generated by the "no change" model of 4.2.1. Since using 200 points in calculating MSE proved to be realisation dependent we will use,

$$\text{MSE} = (1/10) \sum_{r=1}^{10} \text{MSE (realisation } r)$$

which is an average MSE measure and can be expected to have a minimum bias contribution due to realisation peculiarities. Finally to make comparisons easier all MSE values reported will be standardised by reference to the MSE of the standard SSP being set at 100 units (actual value = 0.00295 calculated from Table 4.3).

(ii) $R^{(2)}$ will be measured in terms of z by introducing an outlier at time $t = 21$ on a purely deterministic process.

(iii) $R^{(3)}$ will be illustrated in terms of a graph showing the b_t response of the system to a purely deterministic process with a growth change at time $t = 21$.

(iv) $R^{(4)}$ will similarly be illustrated in terms of a graph showing the m_t response of the system to a purely deterministic process with a step change at time $t = 21$.

Hence the final definition can be simplified as follows:

$$\textit{Performance} = f(\text{MSE}, z, b_t, m_t)$$

which in contrast to any alternative set of purely quantitative criteria makes comparisons of the system's behaviour more meaningful and facilitates subjective interpretation and judgement of the different responses.

CHAPTER 5

Understanding and Controlling the Response of the System

5.1. Introduction

The choice of parameter values in an ICS governs the behaviour of the MSM in the sense that different choices of ICS result in different performance. One ICS may for example produce a very fast growth response $R^{(3)}$, but a bad $R^{(1)}$ response as would be implied by a high MSE value. Another ICS might respond almost instantaneously to step changes of any size, but have an undesirable response to outliers.

Improving the system response to a particular state almost always results in worsening the response to one or more of the other states and a trade off is therefore required. This trade off in performance is however subjective and consequently the aim of this chapter is not to search for a single "optimal" choice of ICS but instead to illusstrate how the different parameters affect performance and how to control the response to a particular state.

In section 5.2 we examine some equivalences between the Π's $\left\{\Pi^{(1)}, \Pi^{(2)}, \Pi^{(3)}, \Pi^{(4)}\right\}$ and the λ's $\left\{\lambda_2, \lambda_3, \lambda_4\right\}$ and show that all the responses of the system can be controlled by either the Π's or the λ's. Varying the Π's however can lead to large instability and therefore we choose to concentrate on the λ's. The effect of the choice of λ's on $R^{(j)}$ is considered in Section 5.3 where the robustness of the system is also tested on different sizes of all modelled discontinuities. In both Section 5.2 and 5.3 the noise variance V_ε is assumed known and therefore $V_{\varepsilon,N}$ is fixed at the V_ε level of 0.0025 which was used in the previous chapter to generate our "no change" data. In Section 5.4 we examine the sensitivity of the system to departures of $V_{\varepsilon,N}$ from V_ε since in practice the noise variance of the process is unknown and hence our estimate of V_ε is expected to be in some error. Finally in Section 5.5 we summarise the conclusions reached in this chapter.

5.2. Interactions between Π's and λ's.

An important characteristic of any model is its ease of control. That is the user should be able to make the model behave in the way he wants. In the context of the MSM our objective is to have a stable system during quiet periods and the ability to control its responses to the different types of discontinuities expected using the minimum possible number of parameters.

In this section we will present evidence to show that all four responses of the system can be controlled by either (i) varying the Π's with the λ's fixed at the their standard SSP values or (ii) varying the λ's with the Π's fixed at their standard SSP values. In addition it is shown that equivalences between Π's and λ's exist in the sense that variations in the Π's only (λ's fixed) possibly reflecting our beliefs about the relative frequencies of occurrence of the different states, produce responses which can be achieved alternatively by variations in the λ's only (Π's fixed).

The reason for the expected equivalences between Π's and λ's is that the responses $R^{(j)}$ depend directly on the posterior probabilities $p_t^{(j)}$ which in turn are calculated in the collapsing process from the posterior transition probabilities $p^{(ij)}$ as given earlier by equation (3.3.9):

$$p^{(ij)} \propto L^{(ij)} \cdot p_{t-1}^{(i)} \cdot \Pi^{(j)}$$

Clearly variations in $p^{(ij)}$ due to $\Pi^{(j)}$ variations only, could alternatively be achieved using the standard $\Pi^{(j)}$ values but altering the λ's which affect the likelihoods, $L^{(ij)}$. These likelihoods can be seen from (3.3.7) and (3.3.8) to depend on the variances $V_\varepsilon^{(j)}$, $V_\mu^{(j)}$. $V_\beta^{(j)}$ which model the different states in terms of the λ's as shown earlier in Table 3.1 of Section 3.4.4.

Let us first consider some extreme cases of $\Pi - \lambda$ equivalences with reference to Table 5.1 below summarising the way the MSM models the environment likely to affect the data series in question in terms of Π's, λ's and $V_{\varepsilon,N}$:

	STATE 1 "no change" j=1	STATE 2 "outlier" j=2	STATE 3 "growth change" j=3	STATE 4 "step change" j=4
$\Pi^{(j)}$	$\Pi^{(1)}$	$\Pi^{(2)}$	$\Pi^{(3)}$	$\Pi^{(4)}$
$V_{\varepsilon}^{(j)}$	$V_{\varepsilon,N}$	$\lambda_2 V_{\varepsilon,N}$	$V_{\varepsilon,N}$	$V_{\varepsilon,N}$
$V_{\mu}^{(j)}$	0	0	0	$\lambda_4 V_{\varepsilon,N}$
$V_{\beta}^{(j)}$	0	0	$\lambda_3 V_{\varepsilon,N}$	0

TABLE 5.1

If $\Pi^{(2)} = 0$ then the system will not recognise outliers since the probability of its occurrence is zero. On the other hand, if $\lambda_2 = 1$ then again outliers can not be recognised since states 1 and 2 now have identical variances (from Table 5.1 with $\lambda_2 = 1$):

$$\left.\begin{array}{lcl} V_{\varepsilon}^{(1)} & = & V_{\varepsilon}^{(2)} \\ V_{\mu}^{(1)} & = & V_{\mu}^{(2)} \\ V_{\beta}^{(1)} & = & V_{\beta}^{(2)} \end{array}\right\}$$

and therefore outliers are no longer modelled.

Similarly $\Pi^{(3)} = 0$ or $\lambda_3 = 0$ would remove the ability to model growth changes while $\Pi^{(4)} = 0$ or $\lambda_4 = 0$ would remove the ability to model step changes. Summarising, SSP choices 1, 2, 3 in the following table, are effectively equivalent to choices 4, 5, 6 respectively.

SSP choice	λ_2	λ_3	λ_4	$\Pi^{(1)}$	$\Pi^{(2)}$	$\Pi^{(3)}$	$\Pi^{(4)}$
Standard SSP	101	1	100	.900	.094	.003	.003
1	standard λ's			.994	0	.003	.003
2				.903	.094	0	.003
3				.903	.094	.003	0
4	1	1	100	standard Π's			
5	101	0	100				
6	101	1	0				

TABLE 5.2

Some extreme cases of equivalencebetween Π's and λ's

In contrast to the extreme cases just considered, finding equivalences between Π's and λ's requires some search until the particular response of our interest is nearly identical. We will now give some further examples to illustrate that we can control every possible response $R^{(j)}$ by varying either the λ's or the Π's from their standard SSP values. The approach followed is to consider the different $R^{(j)}$ responses in turn, for each of a number of SSP's with different Π's but fixed λ's, and to present some other SSP's producing equivalent response but now with Π's fixed and λ's different from their standard SSP values. The search for these equivalences is not described in any way since our only purpose is to illustrate that equivalences do exist. The implication is then that it is not necessary to use all the seven parameters of the SSP in order to control the responses of the system but either the Π's or the λ's are enough.

5.2.1. Controlling $R^{(1)}$

$R^{(1)}$ measures the stability of the system during a "no change" period and one possible way of controlling it, is by varying $\Pi^{(1)}$. Clearly if $\Pi^{(1)}$ is increased then less weight is going to be given to states 2, 3, 4 and therefore greater stability is expected since MSE is calculated over a "no change" period without outliers, growth or step changes. SSP choices 1 and 2 of Table 5.3 illustrate these expectations while choices 3 and 4 have been found to produce equivalent $R^{(1)}$ response.

SSP choice	λ_2	λ_3	λ_4	$\Pi^{(1)}$	$\Pi^{(2)}$	$\Pi^{(3)}$	$\Pi^{(4)}$	MSE
Standard SSP	101	1	100	.9	.094	.003	.003	100
SSP(1)	standard λ's			.99	.0094	.0003	.0003	94
SSP(2)				.9999	.000094	.000003	.000003	90
SSP(3)	101	.1	100	standard Π's				94
SSP(4)	101	.01	100					90

TABLE 5.3
System $R^{(1)}$ response

The reason why SSP(3) and SSP(4) produce equivalent MSE to SSP(1) and SSP(2) respectively is because as λ_3 becomes smaller, the growth change state is making a smaller contribution to the forecasts thus increasing stability.

5.2.2. Controlling $R^{(2)}$

One possible way to affect $R^{(2)}$ using the Π's is by reducing $\Pi^{(2)}$ thus making it more difficult to recognise outliers. Of course a further consequence of this is that the system will become more sensitive to step changes and indeed (as shown in Section 4.1 and Figure 4.1) will respond to the outlier as a step change. SSP choices 1, 2, 3 of Table 5.4 illustrate this point while choices 4, 5, 6 produce equivalent $R^{(2)}$ response by reducing λ_2 and therefore causing an effect similar to that when reducing $\Pi^{(2)}$.

SSP choice	λ_2	λ_3	λ_4	$\Pi^{(1)}$	$\Pi^{(2)}$	$\Pi^{(3)}$	$\Pi^{(4)}$	z
Standard SSP	101	1	100	.9	.094	.003	.003	.3
SSP(1)	standard λ's			.9850	.0090	.0030	.0030	2.5
SSP(2)				.9910	.0030	.0030	.0030	5
SSP(3)				.9937	.0003	.0030	.0030	9
SSP(4)	14	1	100	standard Π's				2.5
SSP(5)	10	1	100					5
SSP(6)	6	1	100					9

TABLE 5.4

$R^{(2)}$ response to a 10σ outlier

5.2.3. Controlling $R^{(3)}$

The speed of response to a growth change is clearly affected by the level of $\Pi^{(3)}$ relative to $\Pi^{(4)}$. If $\Pi^{(3)} > \Pi^{(4)}$, it is expected that small forecast errors will be interpreted as growth changes rather than step changes and therefore an SSP with $(\Pi^{(3)}/\Pi^{(4)})$ ratio higher than another SSP is expected to respond faster to an actual growth change. This is confirmed in Figure 5.1 showing the b_t response of SSP choices 1 and 2 from Table 5.5. SSP(1) has a $\Pi^{(3)}$ to $\Pi^{(4)}$ ratio of 10 while SSP(2) has a ratio of (1/20) and therefore responds slower to the 1σ growth change.

Alternatively the speed of growth response can be affected by the level of λ_3 relative to λ_2 and λ_4. The smaller λ_3 is chosen the slower the growth response since in the limit as λ_3 tends to zero state 3 ("growth change" state) tends to become identical to state 1 ("no change" state) as can be seen from Table 5.1 earlier. SSP choices 3 and 4 have been found to produce almost identical growth response to choices 1 and 2 respectively and the graph of their b_t response is also shown in figure 5.1.

SSP choice	λ_2	λ_3	λ_4	$\Pi^{(1)}$	$\Pi^{(2)}$	$\Pi^{(3)}$	$\Pi^{(4)}$	b_t response	MSE
Standard SSP	101	1	100	.9	.094	.003	.003		100
SSP(1) SSP(2)	standard λ's			.9 .9	.0967 .0601	.0030 .0019	.0003 .0380	shown in Fig.5.1	100 98
SSP(3) SSP(4)	101 101	4 1/4	100 100	standard Π's					100 98

TABLE 5.5

$R^{(3)}$ response to a 1σ growth change

FIGURE 5.1

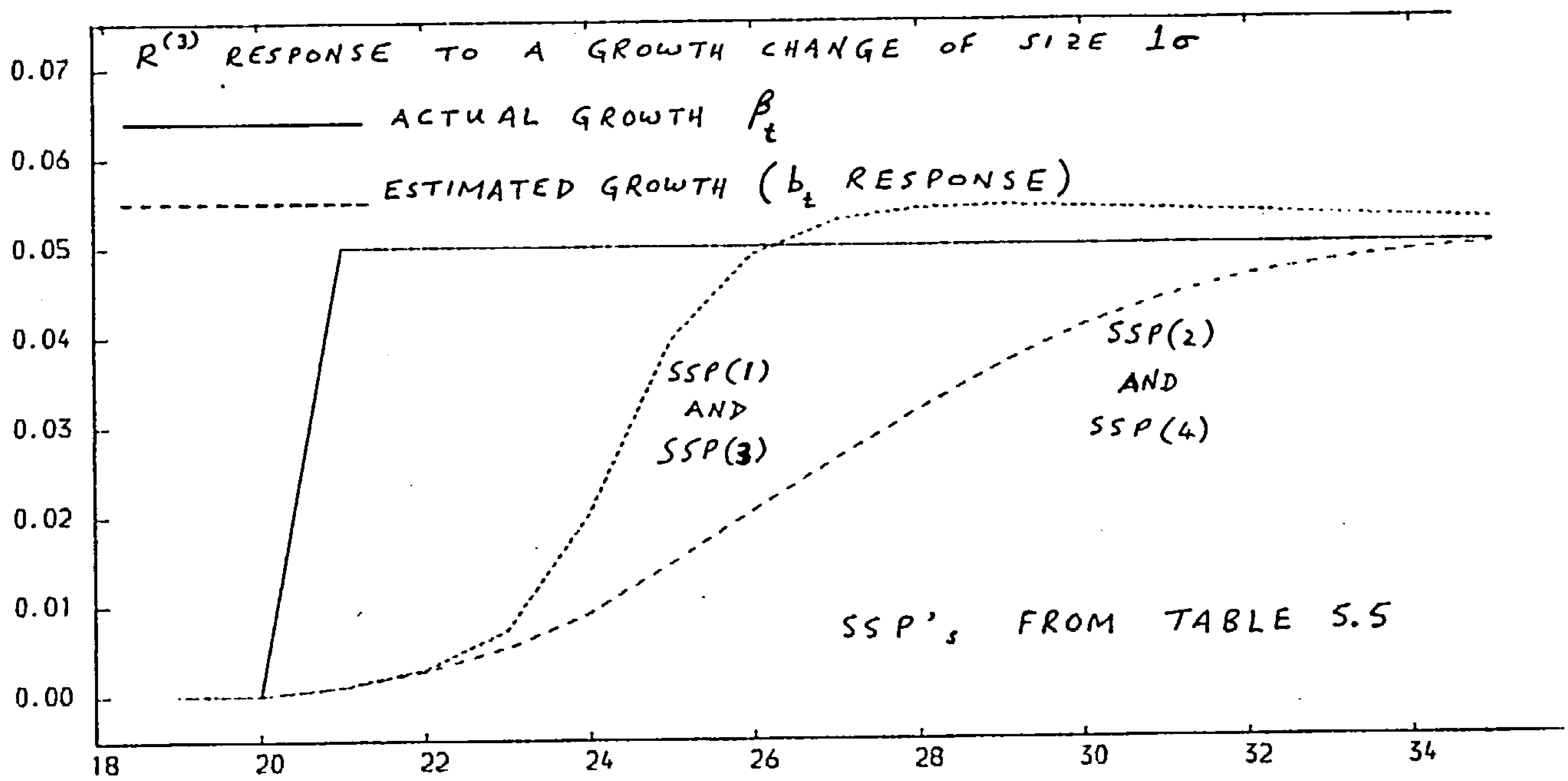
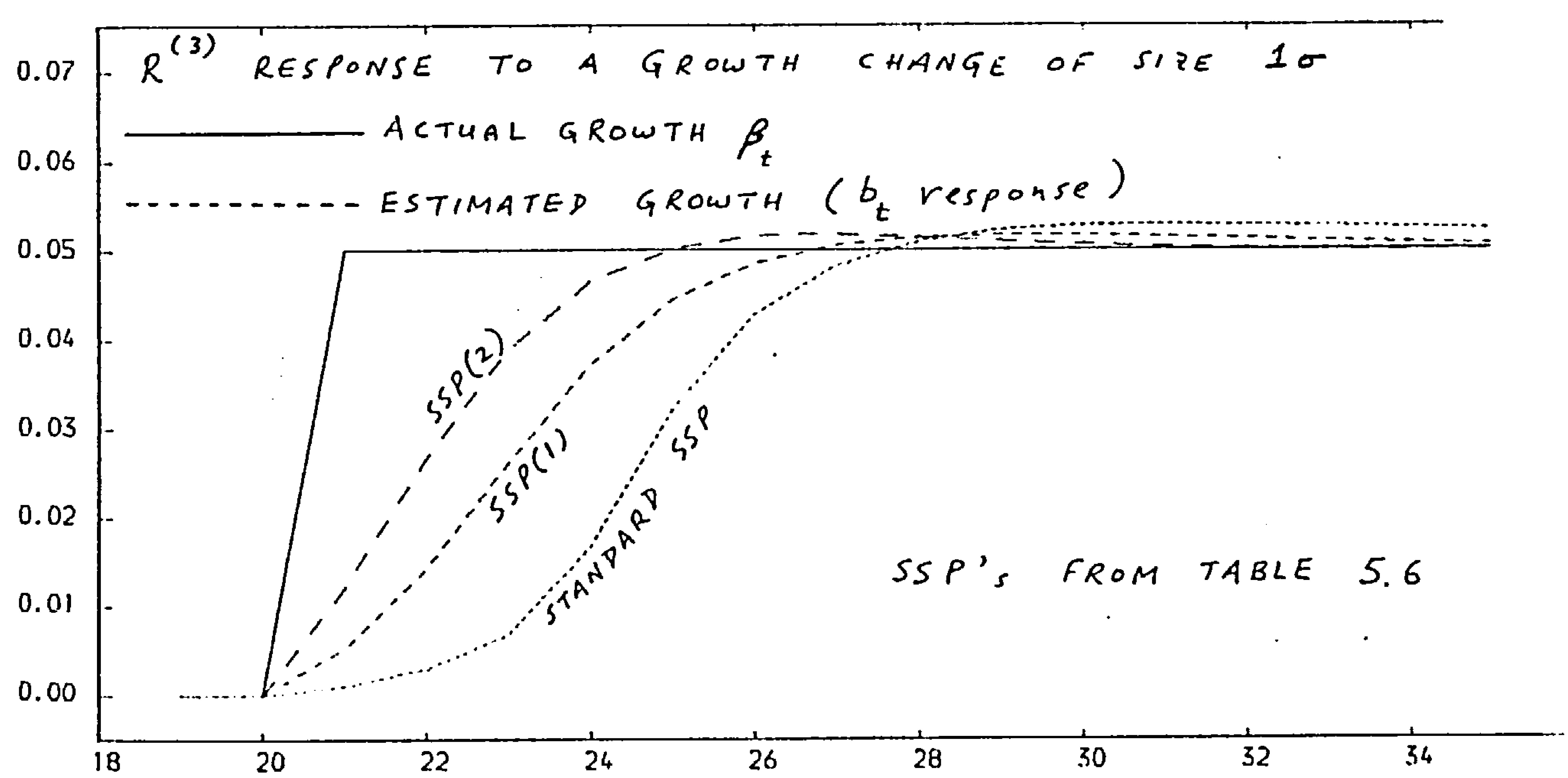


FIGURE 5.2



In an attempt to increase the speed of growth response a common mistake is to change the balance of Π's dramatically by increasing $\Pi^{(3)}$ relative to $\Pi^{(2)}$ and $\Pi^{(4)}$ as well as decreasing $\Pi^{(1)}$. Two such examples are the SSP choices 1 and 2 of Table 5.6 below.

SSP choice	λ_2	λ_3	λ_4	$\Pi^{(1)}$	$\Pi^{(2)}$	$\Pi^{(3)}$	$\Pi^{(4)}$	$R^{(1)}$ MSE	$R^{(3)}$
Standard SSP	101	1	100	.9	.094	.003	.003	100	b_t response to a 1σ growth change is shown in Figure 5.2
SSP(1)	standard λ's			.9	.029	.068	.003	127	
SSP(2)				.7	.029	.268	.003	160	

TABLE 5.6

Figure 5.2 shows that the standard SSP speed of growth response is improved by SSP(1) and SSP(2) successively but this is achieved at the expense of large instability during "no change" periods, since from Table 5.6 we can see that SSP(1) and SSP(2) have a MSE which is higher than that of the standard SSP by 27% and 60% respectively.

To illustrate further the extent of instability caused by changing the balance of Π's from their standard SSP values, we can compare the MSE of SSP(1) and SSP(2) with the MSE of EWR with different discount factors w. Comparisons can also be made with Holt's system since as described in Section 2.4 an EWR with discount factor w is equivalent to Holt's system with smoothing constants $A_1 = 1-w^2$ and

$A_2 = (1-w)^2$.

EWR discount factor w	Holt's equivalent smoothing constants A_1	A_2	$R^{(1)}$ response MSE
.45	.80	.302	189
.50	.75	.250	173
.55	.70	.203	159
.60	.64	.160	147
.65	.58	.123	135
.70	.51	.090	125
.75	.44	.063	116
.80	.36	.040	108
.85	.28	.023	101
.90	.19	.010	94
.95	.10	.003	88

TABLE 5.7

We can see that the standard SSP, SSP(1) and SSP(2) from Table 5.6 are approximately equivalent in MSE to an EWR with $w = .85$, .70 and .55 respectively. For most industrial applications of linear growth models however, the optimal value of w is of the order of .85 or larger. Smaller values of w imply that so little weight is given to historical data that an explanatory model ought to be used instead of a time series model such as EWR, Holt's, MSM etc. A value of $w = .70$ for example is equivalent to Holt's smoothing constants of $A_1 = .5$ and $A_2 = .09$ which is clearly a very unstable system.

Consequently we are interested in data which apart from discontinuities could be followed by an EWR with $w = .85$ or greater. In terms of the response $R^{(1)}$ this necessitates an MSE of 100 or less.

In terms of the SSP's in Table 5.6 this means that the Π's of SSP(1) and SSP(2) produce unacceptably unstable systems and therefore improvements in growth response should not be sought by spoiling the balance of Π's in the standard SSP. The choice of standard Π's offers a good trade off in the conflicting requirements of fast growth response and stability, since it allows good growth response by varying the λ's only and at the same time guarantees stability over all λ values as will be shown in Section 5.3. The speed of growth response offered by the standard SSP can be appreciated by a comparison to the growth response of EWR with $w = .85$ which has the same MSE. The two systems are therefore equivalent in stability during "no change" periods but the MSM produces a much faster growth response than EWR as shown in Figure 5.3.

5.2.4. Controlling $R^{(4)}$

Consider again SSP(1) and SSP(2) from Table 5.5 with $(\Pi^{(3)}/\Pi^{(4)})$ ratios of 10 and (1/20) respectively. In Section 5.2.3. we explained why SSP(1) responds to a growth change faster than SSP(2) as shown in Figure 5.1. However a comparison of their $R^{(4)}$ response (in terms of m_t given a 5σ step change and illustrated in Figure 5.4) indicates a faster response by SSP(2). SSP(1) delays response to a step by one time period and also underestimates the true level slightly over the next few time periods, with the implication that part of the step change has been wrongly interpreted as a growth change thus introducing a small bias into m_t.

FIGURE 5.3

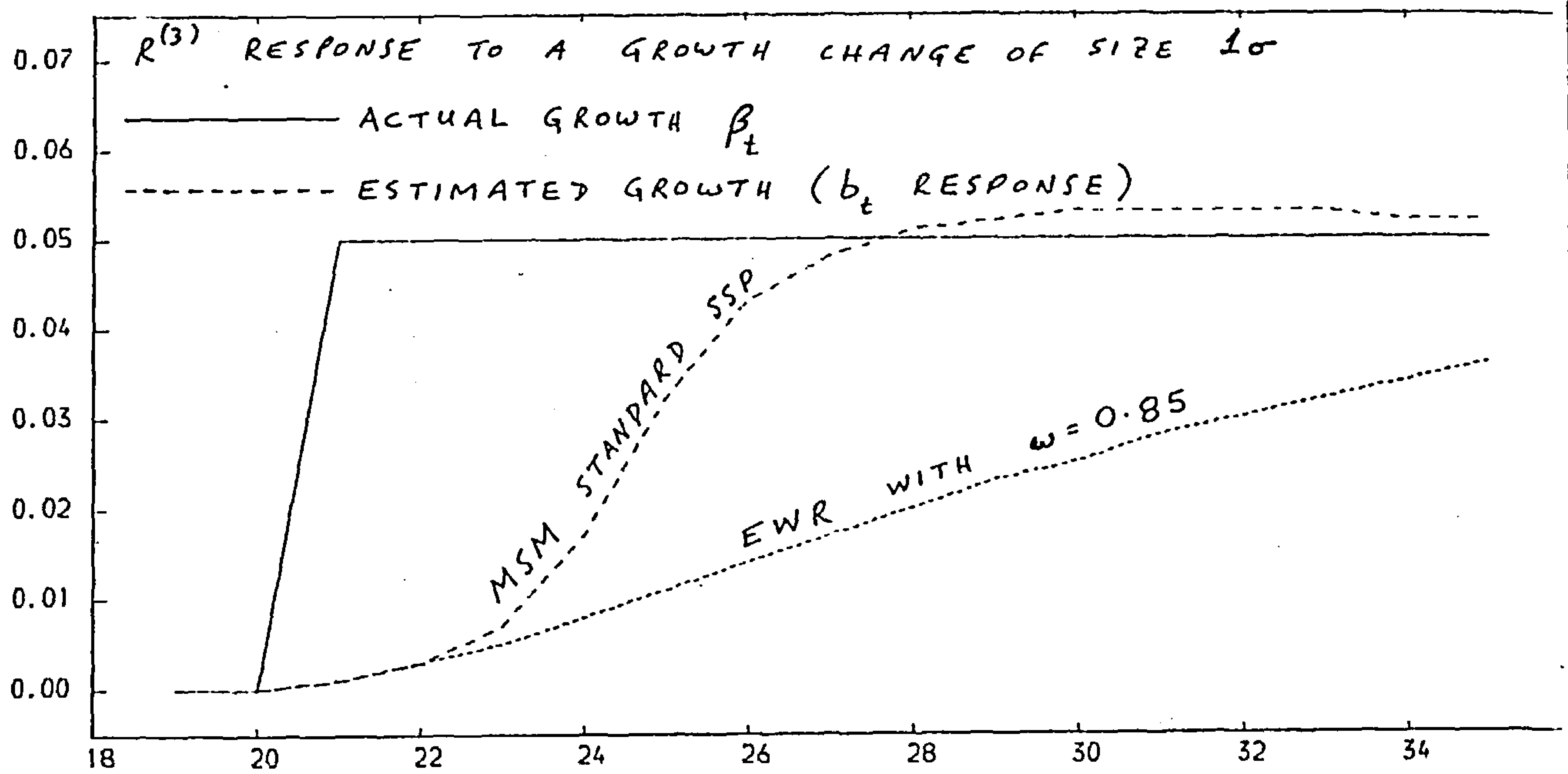
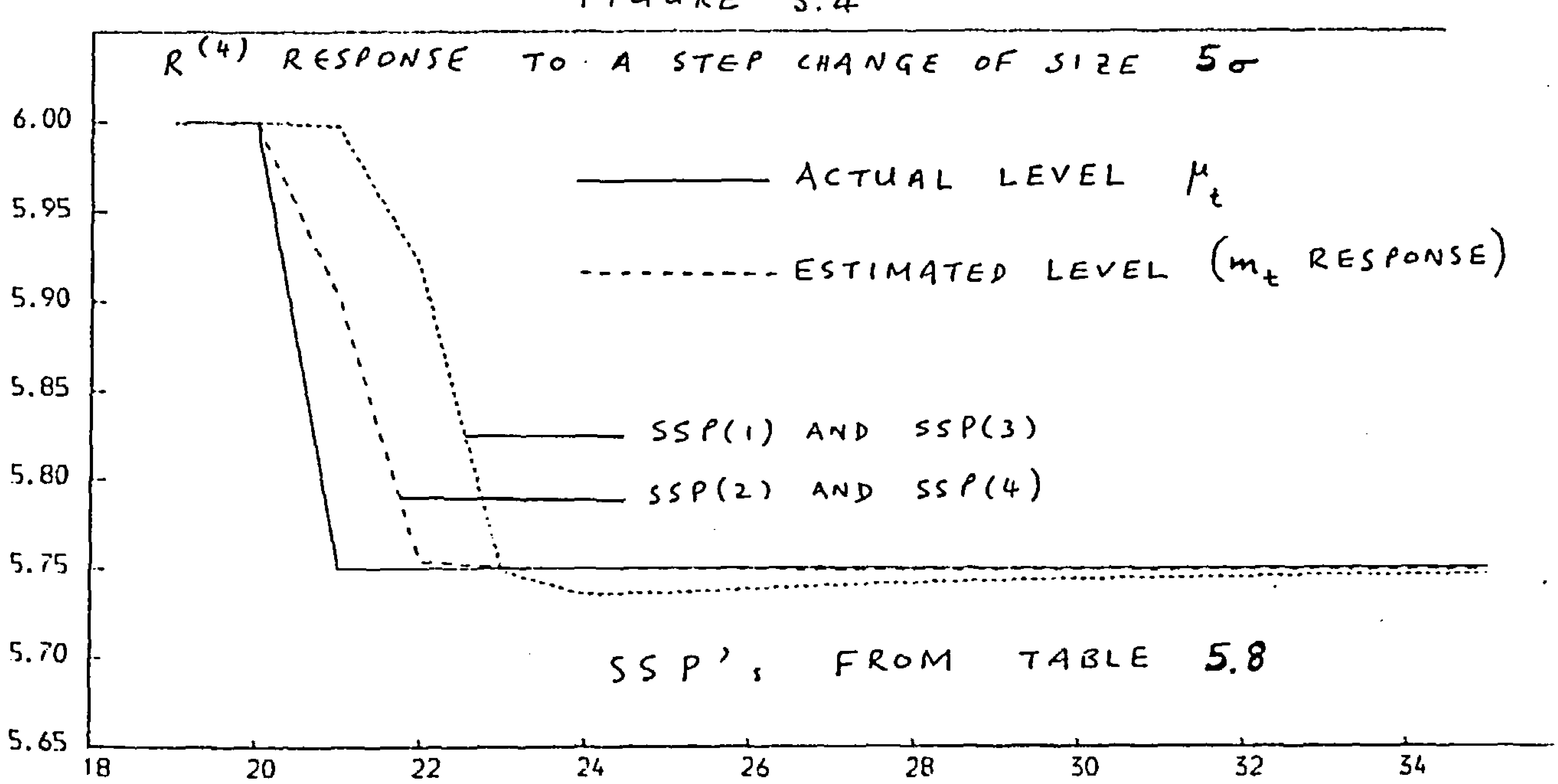


FIGURE 5.4



SSP(3) and SSP(4) in Table 5.8 below, have been found to produce almost identical $R^{(4)}$ responses to SSP(1) and SSP(2) respectively and the graph of their m_t response is also given in figure 5.4.

	λ_2 λ_3 λ_4	$\Pi^{(1)}$ $\Pi^{(2)}$ $\Pi^{(3)}$ $\Pi^{(4)}$	m_t response	MSE
Standard SSP	101 1 100	.9 .094 .003 .003		100
SSP(1) SSP(2)	standard λ's	.9 .0967 .0030 .0003 .9 .0601 .0019 .0380	shown in Fig. 5.4	99 101
SSP(3) SSP(4)	101 1 2 2.5 2.5 100	standard Π's		99 101

TABLE 5.8

Equivalences in $R^{(4)}$ response to a 5σ step change

5.2.5. Concluding remarks

It has been shown that either the λ's or the Π's can be varied from their standard SSP values in order to control any of the four system responses. It was also shown that equivalences between parameter Π's and λ's exist and further that some variations in the Π's can lead to large instability. Hence a good way of understanding the behaviour and controlling the response of the system is through the λ's only. The MSM then becomes a four parameter system requiring values for λ_2, λ_3, λ_4 and an estimate of $V_\varepsilon(V_{\varepsilon,N})$, with the Π's fixed at their standard SSP values.

5.3 Relationship between λ's and $R^{(j)}$

The sensitivity analysis performed in this section is aimed at testing the extent to which the choice of $\underline{\lambda} = \left\{\lambda_2, \lambda_3, \lambda_4\right\}$ is critical to the system's performance and examining the effect on the performance of different sizes of discontinuities. The noise variance is assumed to be known (i.e. $V_{\varepsilon,N} = V_{\varepsilon}$) and the Π's are fixed at their standard SSP values. Three sizes of the different discontinuities are used as follows:

	SIZE		
OUTLIER	4σ	10σ	20σ
GROWTH CHANGE	$(\frac{1}{2})\sigma$	1σ	1.5σ
STEP CHANGE	2σ	5σ	10σ

This range is thought to cover most discontinuities expected in real data but even if one is interested in a size outside this range, the results presented can be used to extrapolate.

Presentation of the results is a problem of great difficulty since for every input $\underline{\lambda}$ set, we have defined ten associated measures of performance:

$$\underline{\lambda} = \left\{\lambda_2, \lambda_3, \lambda_4\right\} \xrightarrow{\text{Input}} \boxed{\text{MSM system}} \xrightarrow[R^{(j)}]{\text{Output}} \begin{cases} R^{(1)} : \text{MSE} \\ R^{(2)} : z \text{ response to outliers of sizes } 4\sigma,\ 10\sigma \text{ and } 20\sigma \\ R^{(3)} : \text{graph of the system growth } b_t \text{ associated with growth changes of size } (\tfrac{1}{2})\sigma,\ 1\sigma \text{ and } 1.5\sigma. \\ R^{(4)} : \text{graph of the system level } m_t \text{ associated with step changes of size } 2\sigma,\ 5\sigma \text{ and } 10\sigma \end{cases}$$

Consider for example the standard $\underline{\lambda}$ set which will be referred to from now on as $\underline{\lambda}_s$,

$$\underline{\lambda}_s = \left\{101,\ 1,\ 100\right\}$$

The ten responses associated with $\underline{\lambda}_s$ are shown in Table 5.9 and Figures 5.5, 5.6 and they will be used as a benchmark against which all other responses are to be compared.

$R^{(1)}$	$R^{(2)}$		
	z response to outliers of size:		
MSE	4σ	10σ	20σ
100	.3	.3	.7

TABLE 5.9

$R^{(1)}$ and $R^{(2)}$ associated with $\underline{\lambda}_s$

The small z values indicate that the system is unaffected by even large outliers and from Figures 5.5 and 5.6 we can see that small growth and step changes are harder to detect but as they increase in size the speed and accuracy of the system's responses also increases. From Figure 5.5 for example it can be seen that the average number of time periods required to cover 90% of the growth discontinuity is approximately 11, 7 and 5 for growth changes of $(\frac{1}{2})\sigma$, 1σ and 1.5σ respectively. Similarly from Figure 5.6 we can see that for steps of size 5σ or larger the system level is hardly affected at time $t = 21$ (since the first large forecast error is treated as an outlier) but then responds almost instantaneously. For smaller step changes however such as 2σ in size, the system level responds slowly since the small forecast errors realised are easily interpreted as "no change" by the model characterising state 1 (defined in Section 3.2 as $M_t^{(1)}$). This model responds to the small step change by slowly changing its growth component (defined in Section 3.3 as $b_t^{(1)}$).

R(3) RESPONSE
ACTUAL GROWTH βt
ESTIMATED GROWTH (bt RESPONSE)
½ σ GROWTH CHANGE
1 σ GROWTH CHANGE
1·5 σ GROWTH CHANGE
λs

FIGURE 5.5

$R^{(4)}$ RESPONSE

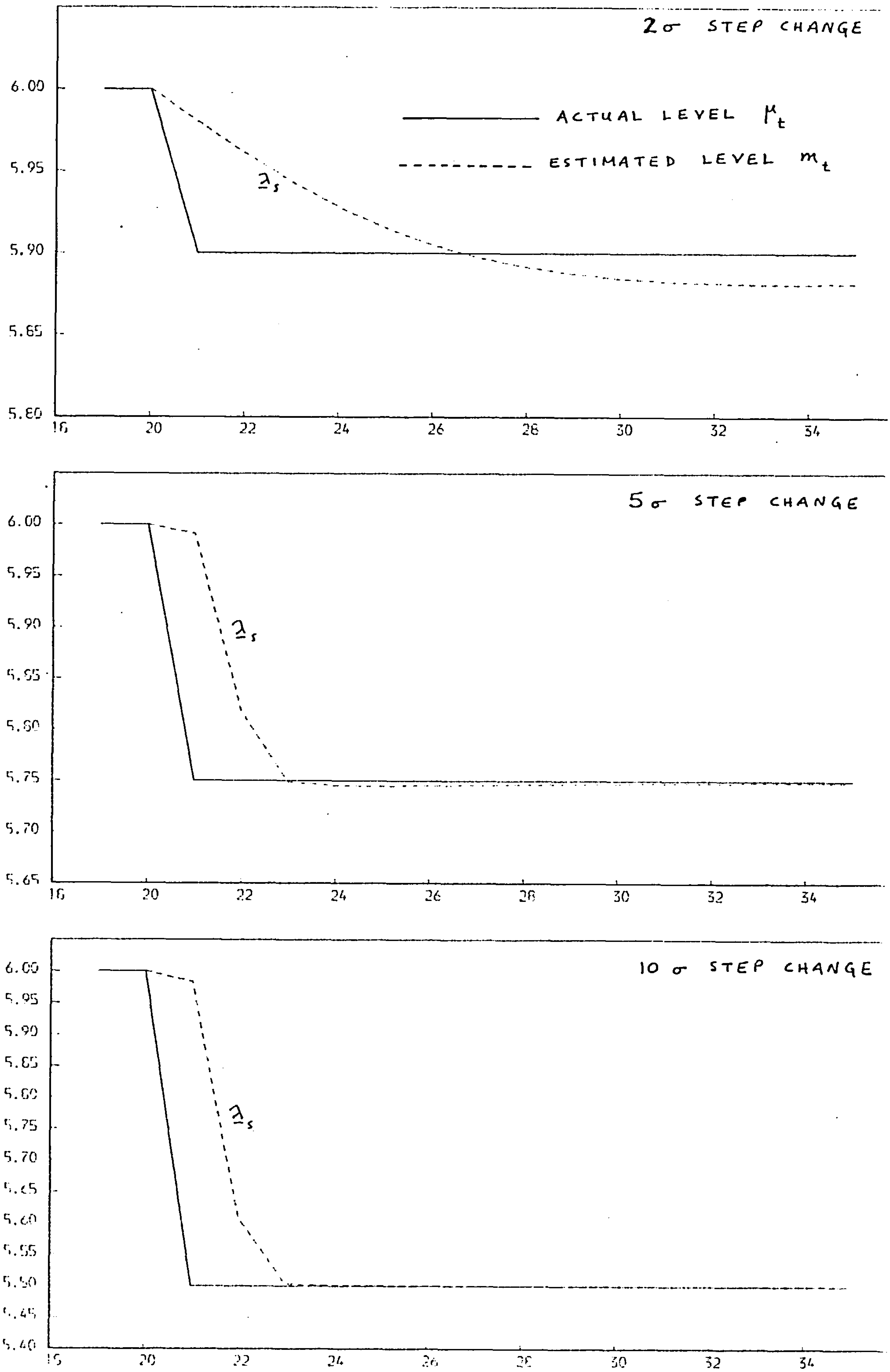


FIGURE 5.6

Depending on the nature of the process under consideration one or more of the states modelling discontinuities may not occur and consequently the system response to these states is irrelevant. It is expected that choosing the λ's so as to exclude these states may improve the responses to the relevant states since by modelling fewer states the uncertainty in interpreting a change is reduced. This still leaves the problem as to the magnitude of the λ's corresponding to the remaining states and the response to these states. We therefore start by examining combinations of states 2 at a time and then investigate possible interactions between λ's by examining states 3 at a time and finally all four. Recalling that states 1, 2, 3, 4 correspond to "no change", "outlier", "growth change" and "step change" respectively then we can use the following notation for the different situations:

2-state models	MSM(1,2) :	MSM	modelling	states	1 and 2 only
	MSM(1,3) :	"	"	"	1 and 3 "
	MSM(1,4) :	"	"	"	1 and 4 "
3-state models	MSM(1,2,3) :	"	"	"	1, 2 and 3 only
	MSM(1,2,4) :	"	"	"	1, 2 and 4 "
	MSM(1,3,4) :	"	"	"	1, 3 and 4 "
4-state model	MSM(1,2,3,4):	"	"	"	all four states

5.3.1. MSM(1,2)

Here we are not interested in $R^{(3)}$ and $R^{(4)}$ and therefore the "growth change" and "step change" states are excluded by setting $\lambda_3 = 0$ and $\lambda_4 = 0$ which reduces states 3 and 4 to "no change" states as can be seen from Table 5.1 in the beginning of this chapter. The MSM is therefore a 2-state model with a single parameter λ_2 which controls the response to outliers. The results for different values of λ_2 are summarised in Table 5.10 below.

	λ_2	λ_3	λ_4	$R^{(2)}$ z response to outliers of size 4σ	10σ	20σ	$R^{(1)}$ MSE
$\underline{\lambda}_s$:	101	1	100	.3	.3	.7	100
	1	0	0	.8	1.9	3.8	89
	2			.6	1.1	2.1	89
	4			.3	.6	1.1	88
	10			.1	.2	.5	88
	25			.1	.1	.2	88
	50			.1	.0	.1	88
	101			.1	.0	.0	88
	200			.1	.0	.0	88
	400			.2	.0	.0	88
	1000			.2	.0	.0	89
	$(10)^{20}$			.8	.0	.0	89

TABLE 5.10

Responses of MSM(1,2)

It can be seen that the exclusion of states 3 and 4 reduces the MSE significiantly, from 100 of $\underline{\lambda}_s$ to approximately 88 irrespective of the choice of λ_2. When $\lambda_2 = 1$ the outlier state is also excluded and as a result the system responds in an undesirable way (too large z) to outliers. As λ_2 becomes larger the z response improves and for $\lambda_2 \geqslant 10$ the system produces a marginally better z response than that of $\underline{\lambda}_s$. A value of λ_2 between 101 and 200 seems the best for the range of outlier sizes considered, but even larger values of λ_2 are not critical to either $R^{(1)}$ nor $R^{(2)}$.

5.3.2. MSM(1,3)

We now exclude states 2 and 4 by setting $\lambda_2 = 1$ and $\lambda_4 = 0$ since the only responses of interest in this model are $R^{(1)}$ and $R^{(3)}$, that is the stability of the system and its speed of response to growth changes. The results obtained for various values of λ_3 are shown in Table 5.11 and Figures 5.7a and 5.7b.

	λ_2	λ_3	λ_4	$R^{(1)}$ MSE	$R^{(3)}$ b_t response
λ_3 :	101	1	100	100	
	1	1/10	0	97	
		1/4		99	see
		1/2		101	figures
		1		102	5.7a
		4		101	and
		16		99	5.7b
		64		97	
		256		95	

TABLE 5.11

Responses of MSM(1,3)

From figures 5.7a and 5.7b we can see that values of λ_3 in the range 1 to 4 produce the fastest growth response and the MSE (from Table 5.11) is at its maximum. Values of λ_3 smaller than 1, produce more stable response during quiet periods (i.e. lower MSE) but reduce the speed of growth response significantly. A value of $\lambda_3 = 1/10$ for example has a MSE approximately 5% lower than $\lambda_3 = 1$ but requires approximately 5 extra time periods in order to cover 90% of the growth discontinuities. Values of λ_3 greater than 4 also produce a more stable system at the expense of marginally slower growth response and slightly more overshooting, but are not critical as can be the case when $\lambda_3 << 1$.

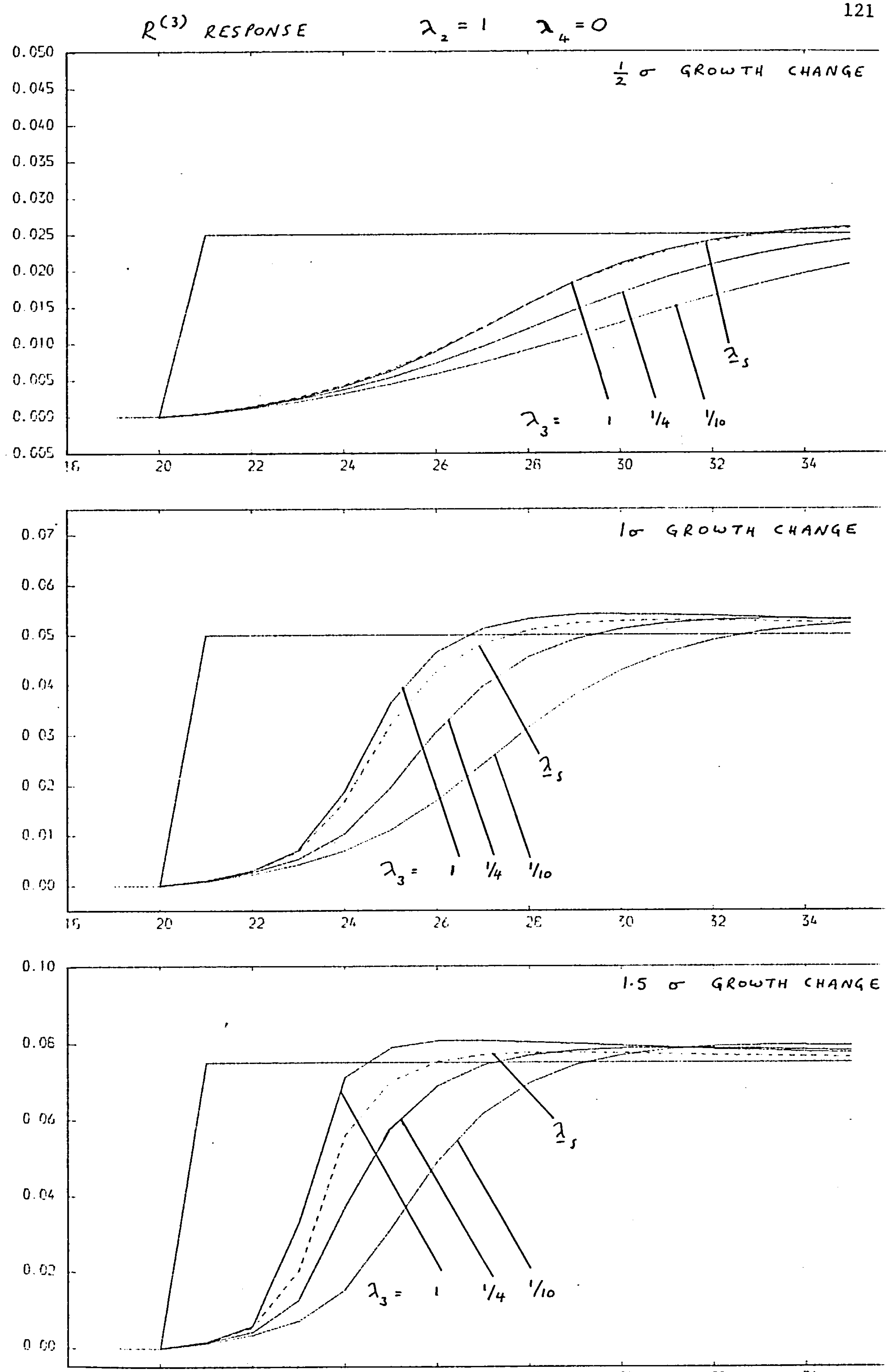

$R^{(3)}$ RESPONSE $\lambda_2 = 1$ $\lambda_4 = 0$
$\frac{1}{2}\sigma$ GROWTH CHANGE
$\lambda_3 =$ 1 1/4 1/10
λ_s
1σ GROWTH CHANGE
1.5 σ GROWTH CHANGE

FIGURE 5.7.a

$R^{(3)}$ RESPONSE $\lambda_2 = 1$ $\lambda_4 = 0$
$\frac{1}{2}\sigma$ GROWTH CHANGE
λ_s
λ_3 = 4 16 64 256
1σ GROWTH CHANGE
λ_s
λ_3 =
256
64
16
4
1.5 σ GROWTH CHANGE
λ_s
λ_3 =
256
64
16
4

FIGURE 5.7.6

5.3.3. MSM(1,4)

We are now interested in $R^{(1)}$ and $R^{(4)}$ only and we therefore exclude states 2 and 3 by setting $\lambda_2 = 1$ and $\lambda_3 = 0$. The results from a large range of λ_4 values are given in Table 5.12 and Figures 5.8a and 5.8b.

	λ_2	λ_3	λ_4	$R^{(1)}$ MSE	$R^{(4)}$ m_t response
$\underline{\lambda}_s$:	101	1	100	100	
	1	0	1/2	90	
			1	90	
			4	90	see figures
			10	90	5.8a and
			25	89	5.8b
			50	89	
			100	89	
			200	89	
			400	89	
			1000	89	

TABLE 5.12
Responses of MSM(1,4)

Like in MSM(1,2) the MSE is significantly reduced from the $\underline{\lambda}_s$ level of 100, and is not affected by the value of λ_4. From Figures 5.8a and 5.8b it can be seen that for a 2σ step change, λ_4 does not affect the system level response. This implies that the "step change" model $M_t^{(4)}$, does not recognise such small changes which can easily be

$R^{(4)}$ RESPONSE $\lambda_2 = 1$ $\lambda_3 = 0$

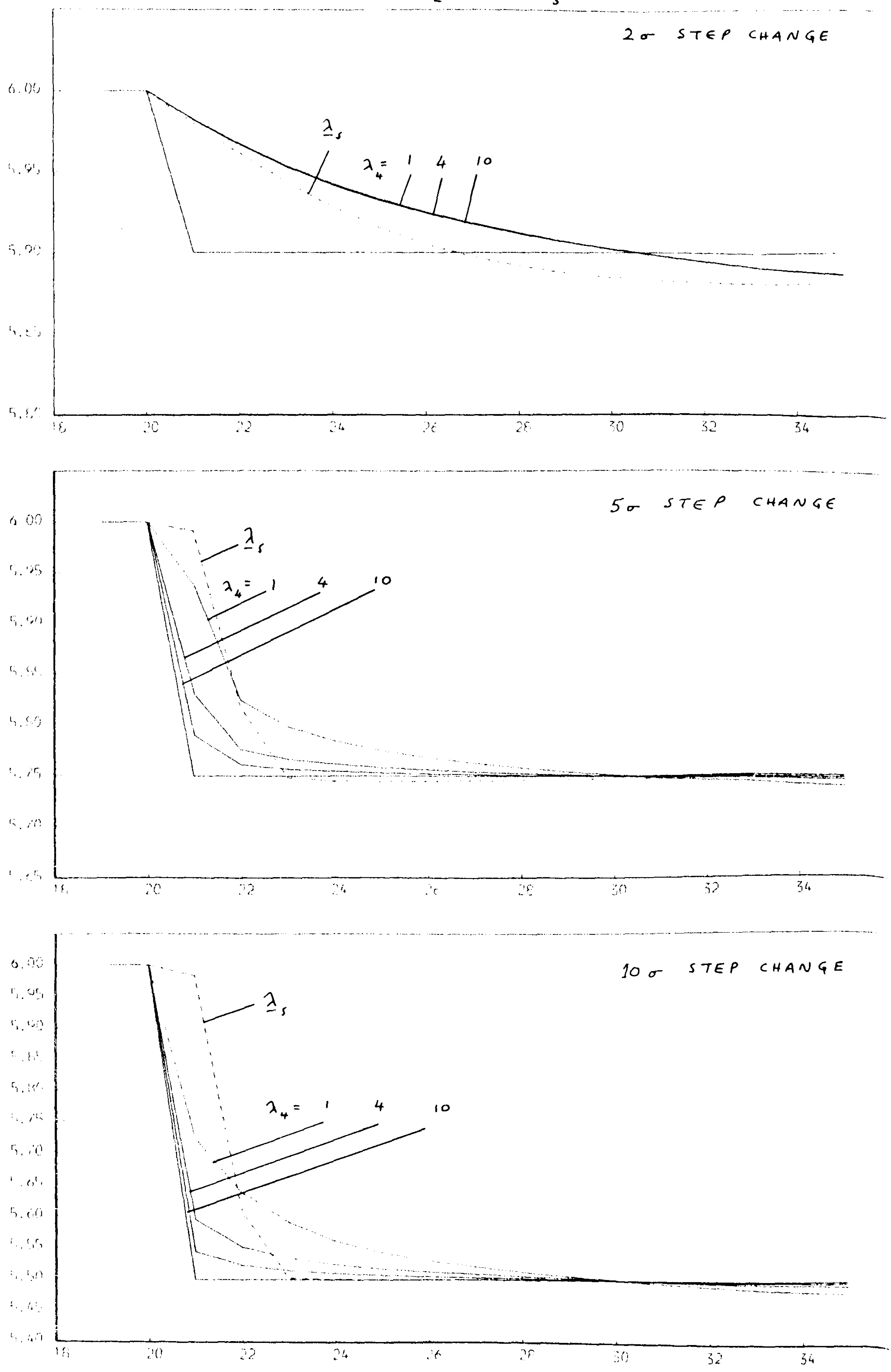


FIGURE 5.8.a


$R^{(4)}$ RESPONSE $\lambda_2 = 1$ $\lambda_3 = 0$
2σ STEP CHANGE
λ_s
$\lambda_4 = 25$ 100 400 1000
6.00
5.95
5.90
5.85
5.80
18 20 22 24 26 28 30 32 34
5σ STEP CHANGE
λ_s
$\lambda_4 =$
1000
400
100
25
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
18 20 22 24 26 28 30 32 34
10 σ STEP CHANGE
λ_s
$\lambda_4 = 25$ 100 400 1000
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
5.60
5.55
5.50
5.45
5.40
18 20 22 24 26 28 30 32 34

FIGURE 5.8.6

followed by the "no change" model $M_t^{(1)}$. For the larger step sizes however the system level response improves as λ_4 increases up to a value of approximately 25 when the response is almost instantaneous. Larger values of λ_4 result in nearly identical response and no further improvement is possible. Comparing the system level response of MSM(1,4) with $\lambda_4 \geq 25$ against that of $\underline{\lambda}_s$ we can see that MSM(1,4) responds to a step change faster mainly because the outlier state is not modelled and therefore given a large discontinuity it is interpreted as a step change almost immediately. We can therefore conclude that for a process where growth changes and outliers are not expected, MSM(1,4) is a better model than MSM with $\underline{\lambda}_s$ since it not only responds faster to step changes but also has a significantly smaller MSE.

5.3.4. MSM(1,2,3)

State 4 is excluded by setting $\lambda_4 = 0$ since the only irrelevant response in this model is $R^{(4)}$. The results from a factorial experiment varying λ_2 and λ_3 are shown in Table 5.13. Values of $\lambda_2 < 10$ and λ_3 outside the range $[1/4, 64]$ have not been included since from MSM(1,2) and MSM(1,3) we have seen that they lead to worse $R^{(2)}$ and $R^{(4)}$ responses respectively.

λ_2	λ_3	λ_4	$R^{(1)}$ MSE	$R^{(2)}$ z response to outliers of size: 4σ	10σ	20σ	$R^{(3)}$ b_t response
$\underline{\lambda}_s$: 101	1	100	100	.3	.3	.7	
10	1/4	0	97	.2	.3	.6	
	1		99	.2	.3	.7	
	4		99	.3	.3	1.3	
	16		97	.4	3.1	36.8	
	64		95	.3	8.7	39.3	see
101	1/4		97	.1	.0	.1	figures
	1		100	.2	.0	.1	5.9a
	4		100	.3	.0	.1	and
	16		98	.5	.2	.1	5.9b
	64		96	.4	.6	.6	
1000	1/4		99	.3	.0	.0	
	1		101	.4	.0	.0	
	4		101	.7	.0	.0	
	16		99	1.0	.3	.0	
	64		96	.8	1.1	.3	

TABLE 5.13

Responses of MSM(1,2,3)

Most of the results shown in Table 5.13 reflect the small interaction between λ_2 and λ_3. However low values of λ_2 combined with high values of λ_3 (e.g. $\lambda_2 = 10$, $\lambda_3 = 64$) result in undesirable response to outliers. In addition it has been found that λ_2 does not significantly affect the growth response of the system irrespective of the value of λ_3. One example for $\lambda_3 = 1$ is given in Figure 5.9a


$R^{(3)}$ RESPONSE $\lambda_3 = 1$ $\lambda_4 = 0$
$\frac{1}{2}\sigma$ GROWTH CHANGE
$\lambda_2 =$ 101
1000
λ_s
0.050
0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
-0.005
16
20
22
24
26
28
30
32
34
1σ GROWTH CHANGE
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
1.5σ GROWTH CHANGE
0.10
0.08
0.06
0.04
0.02
0.00

FIGURE 5.9.a

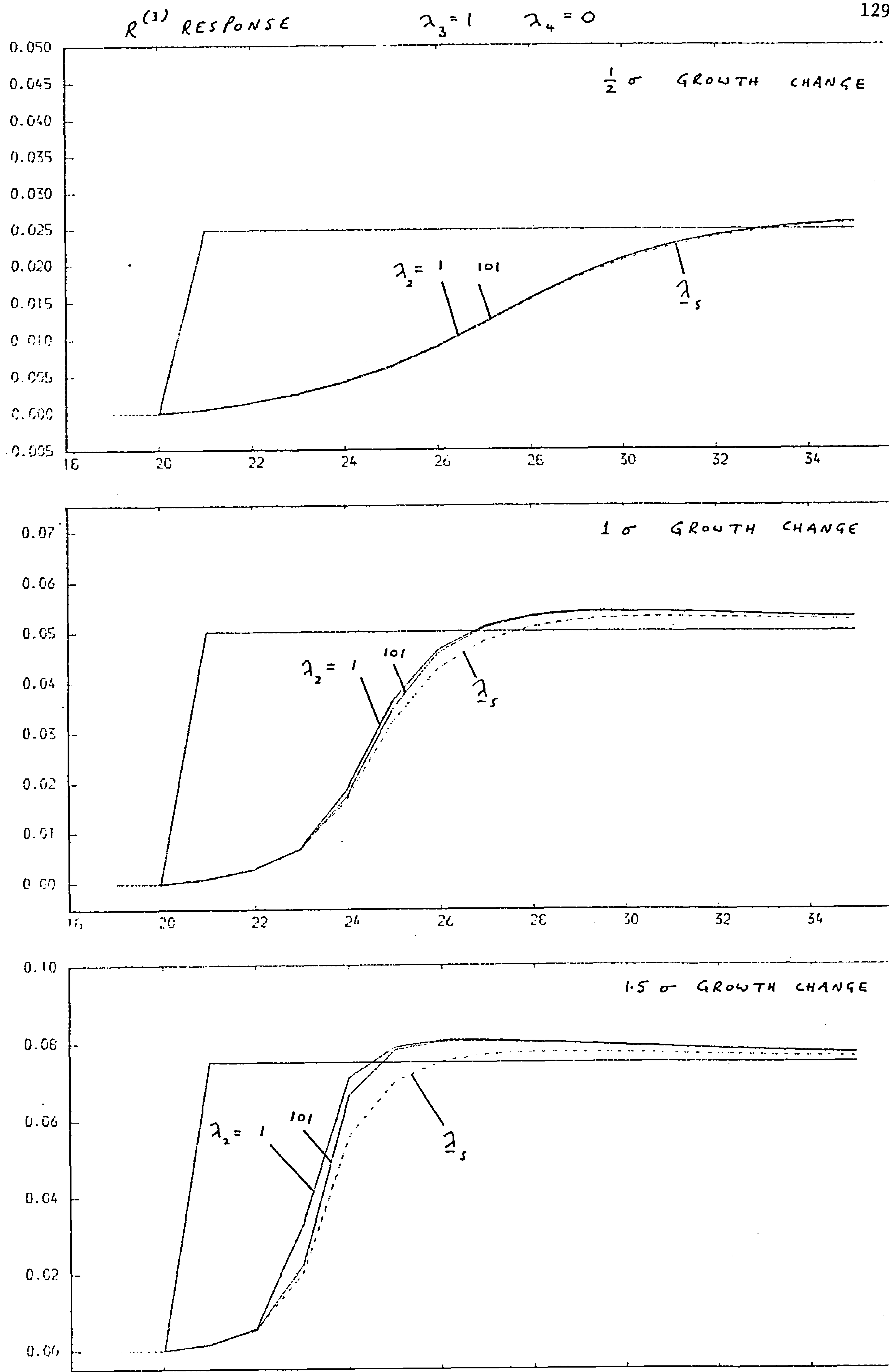

R(3) RESPONSE λ3 = 1 λ4 = 0
½ σ GROWTH CHANGE
λ2 = 1
101
λs
1 σ GROWTH CHANGE
λ2 = 1
101
λs
1.5 σ GROWTH CHANGE
λ2 = 1
101
λs

FIGURE 5.9.b

where $\lambda_2 = 101$ and $\lambda_2 = 1000$ produce nearly identical $R^{(3)}$. This together with the fact that $\lambda_2 = 101$ has marginally smaller MSE than $\lambda_2 = 1000$ over all λ_3 values shown in Table 5.13, suggest that a value of $\lambda_2 = 101$ is the best since it has good $R^{(2)}$ response, minimum MSE over all $\lambda_2 \geq 101$ and does not spoil the growth response $R^{(3)}$.

Given that λ_2 does not affect $R^{(3)}$ it follows that the growth response of the system depends only on λ_3 and it is in fact nearly identical to that produced by MSM(1,3) where λ_2 was fixed at the value of 1 thus excluding state 2. Figure 5.9b illustrates this for the case of $\lambda_3 = 1$. Other graphs of $R^{(3)}$ for different λ_3 are not given since they are almost identical to those given earlier for MSM(1,3) in figures 5.7a and 5.7b.

Comparing Tables 5.13 and 5.11 for different λ_3 we can see that the introduction of λ_2 in MSM(1,2,3) improves the MSE by approximately 2% and therefore including state 2 (thus modelling outliers) increases the stability of the system without spoiling the growth response $R^{(3)}$ as shown above. The surprising implication is that MSM(1,2,3) should be used instead of MSM(1,3) even when outliers are not expected.

5.3.5.MSM(1,2,4)

The relevant responses are now $R^{(1)}$, $R^{(2)}$ and $R^{(4)}$ and therefore state 3 is excluded by setting $\lambda_3 = 0$. The results from a factorial experiment varying λ_2 and λ_4 are shown in Table 5.14

and Figures 5.10a, 5.10b. Values of $\lambda_2 < 10$ and $\lambda_4 < 25$ are not included since they lead to worse $R^{(2)}$ and $R^{(4)}$ responses as was shown in MSM(1,2) and MSM(1,4) respectively.

λ_2	λ_3	λ_4	$R^{(1)}$ MSE	$R^{(2)}$ z response to outliers of size: 4σ	10σ	20σ	$R^{(4)}$ m_t response
$\underline{\lambda}_s$: 101	1	100	100	.3	.3	.7	
10	0	25	88	.3	2.9	19.2	see figures 5.10a and 5.10b
		100	88	.2	4.6	19.8	
		400	88	.2	3.9	20.0	
		1000	88	.2	3.1	20.0	
25		25	88	.2	.4	1.0	
		100	88	.2	.7	17.1	
		400	88	.1	.6	18.6	
		1000	88	.1	.4	18.4	
101		25	88	.3	.2	.1	
		100	88	.2	.3	.7	
		400	88	.2	.2	1.4	
		1000	88	.1	.2	1.2	
400		25	88	.4	.2	.0	
		100	88	.3	.4	.3	
		400	88	.2	.3	.6	
		1000	88	.2	.2	.5	
1000		25	89	.6	.3	.0	
		100	88	.4	.6	.3	
		400	88	.3	.4	.7	
		1000	89	.3	.3	.6	

TABLE 5.14

Responses of MSM(1,2,4)

$R^{(4)}$ RESPONSE $\lambda_2 = 101$ $\lambda_3 = 0$

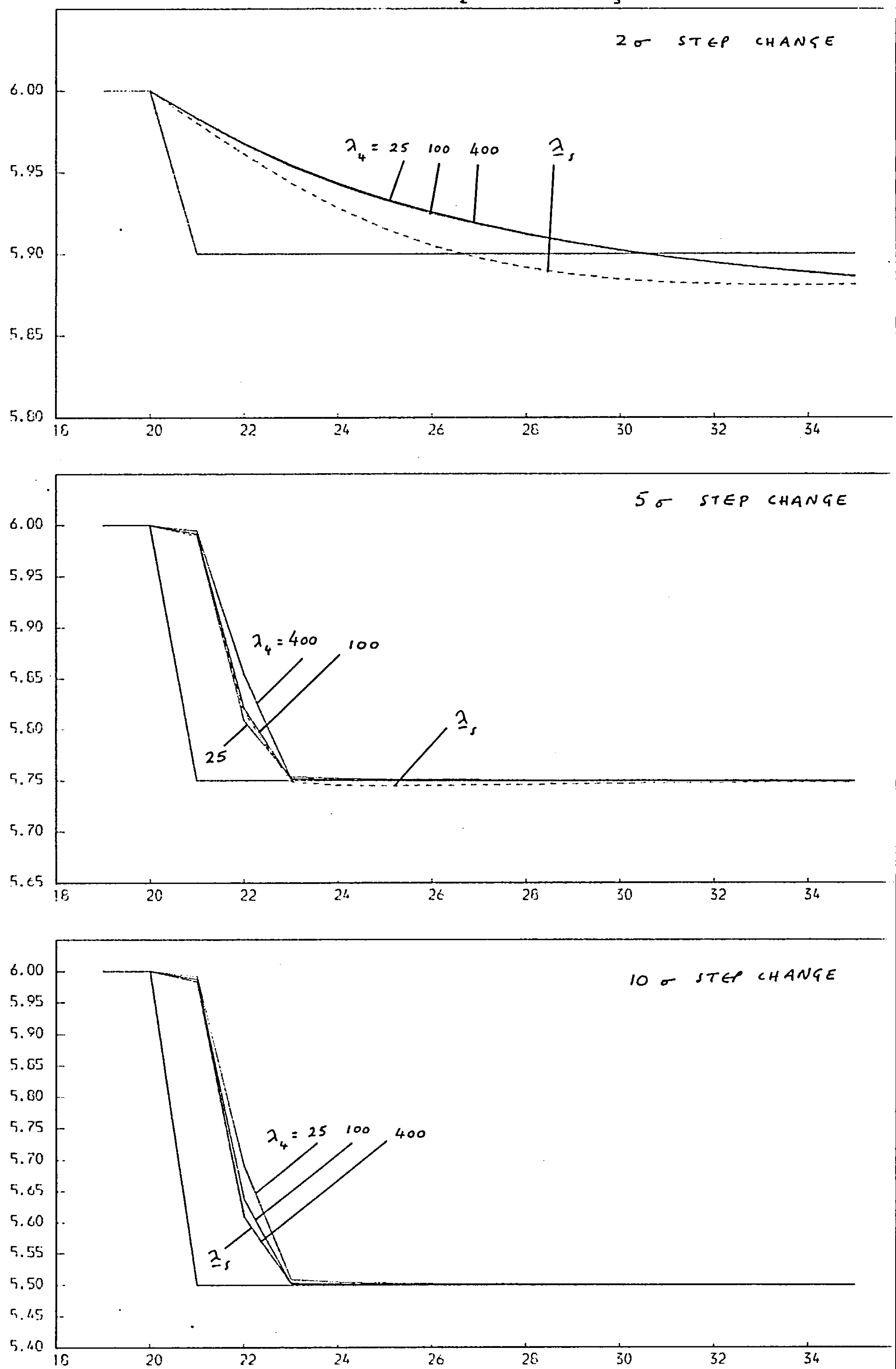


FIGURE 5.10.a

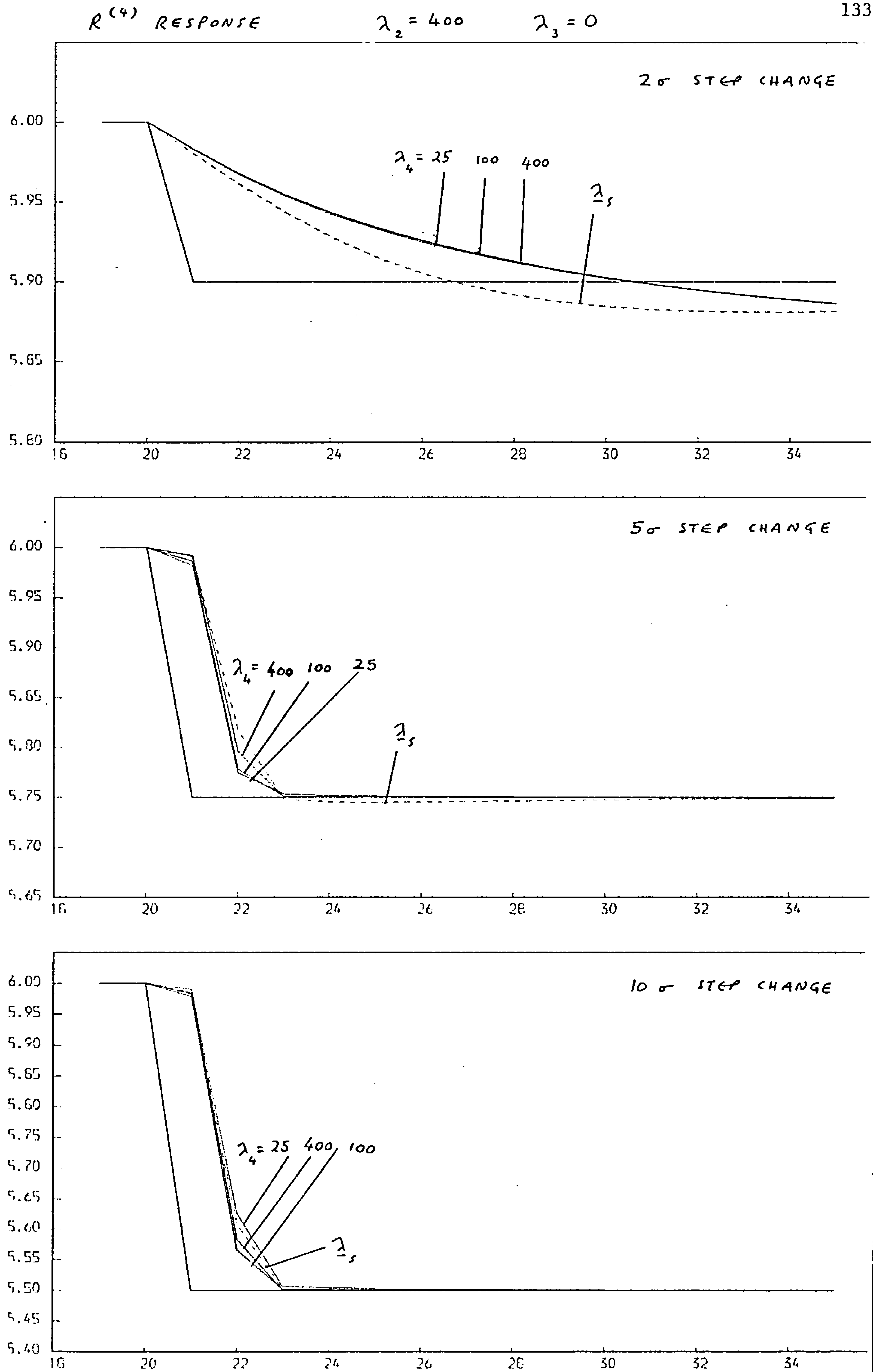

$R^{(4)}$ RESPONSE $\lambda_2 = 400$ $\lambda_3 = 0$
2σ STEP CHANGE
$\lambda_4 = 25$ 100 400
λ_s
6.00
5.95
5.90
5.85
5.80
16 20 22 24 26 28 30 32 34
5σ STEP CHANGE
$\lambda_4 = 400$ 100 25
λ_s
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
16 20 22 24 26 28 30 32 34
10 σ STEP CHANGE
$\lambda_4 = 25$ 400, 100
λ_s
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
5.60
5.55
5.50
5.45
5.40
16 20 22 24 26 28 30 32 34

FIGURE 5.10. b

From Table 5.14 it can be seen that the MSE is not significantly affected by λ_2 and λ_4 and is significantly better than the MSE of $\underline{\lambda}_s$. It can also be seen that for values of $\lambda_2 < 101$, λ_2 interacts with λ_4 producing an undesirable z response especially for large outliers. This is because when λ_2 is too small, large discontinuities are more easily interpreted as step changes rather than outliers. However for $\lambda_2 \geq 101$ the z response is good irrespective of the value of λ_4.

Finally the system level response $R^{(4)}$ is not significantly affected by the relative levels of λ_2 and λ_4 and behaves similarly to $R^{(4)}$ of MSM(1,4) which had λ_4 as the only factor with λ_2 fixed at the value of 1 thus excluding the outlier state. The only effect the introduction of $\lambda_2 \geq 101$ has on $R^{(4)}$ is that the system level does not respond to the first large forecast error at time $t = 21$, since this is now interpreted as an outlier. Figures 5.10a and 5.10b illustrate these points by showing $R^{(4)}$ produced by a number of (λ_2, λ_4) pairs.

5.3.6 MSM(1,3,4)

The only irrelevant response in this model is $R^{(2)}$ and therefore the outlier state is excluded by setting $\lambda_2 = 1$. The results from a factorial experiment varying λ_3 and λ_4 are shown in Table 5.15 and Figures 5.11 and 5.12.

The MSE in Table 5.15 is virtually unaffected by the different values of λ_4 and depends only on λ_3 in an identical way that the

MSE of MSM(1,3) was seen to depend on λ_3 in Table 5.11. The effect of λ_4 in $R^{(3)}$ has also been found to be insignificant and some examples illustrating this, are given in Figures 5.11a, 5.11b and 5.11c. A further comparison of these three figures with Figures 5.7a and 5.7b showing $R^{(3)}$ of MSM(1,3) suggests however that although the introduction of λ_4 in MSM(1,3,4) does not affect the speed of growth response for $\lambda_3 \geq 1$ it significantly reduces the speed of growth response for $\lambda_3 = 1/4$. Values of $\lambda_3 < 1$ are therefore far more critical to $R^{(3)}$ when both step and growth changes are modelled as in MSM(1,3,4) than when only growth changes are modelled as in MSM(1,3).

λ_2	λ_3	λ_4	$R^{(1)}$ MSE	$R^{(3)}$, $R^{(4)}$ b_t and m_t response
$\underline{\lambda}_s$: 101	1	100	100	
1	1/4	10	99	
		25	99	
		100	99	
		400	99	
		1000	99	see
	1	10	102	figures
		25	102	5.11a
		100	102	5.11b
		400	102	5.11c
		1000	102	
	4	10	101	
		25	101	
		100	101	and
		400	101	
		1000	101	
	16	10	99	5.12a
		25	99	5.12b
		100	99	5.12c
		400	99	
		1000	99	
	64	10	97	
		25	97	
		100	97	
		400	97	
		1000	97	

TABLE 5.15
Responses of MSM(1,3,4)


$R^{(3)}$ RESPONSE $\lambda_2 = 1$ $\lambda_3 = 1/4$
$\frac{1}{2}\sigma$ GROWTH CHANGE
λ_5
$\lambda_4 = 25$ 100 400
1σ GROWTH CHANGE
λ_5
$\lambda_4 = 25$ 100 400
1.5σ GROWTH CHANGE
λ_5
$\lambda_4 = 25$ 100 400
18 20 22 24 26 28 30 32 34

FIGURE 5.11.a

$R^{(3)}$ RESPONSE $\lambda_2 = 1$ $\lambda_3 = 1$

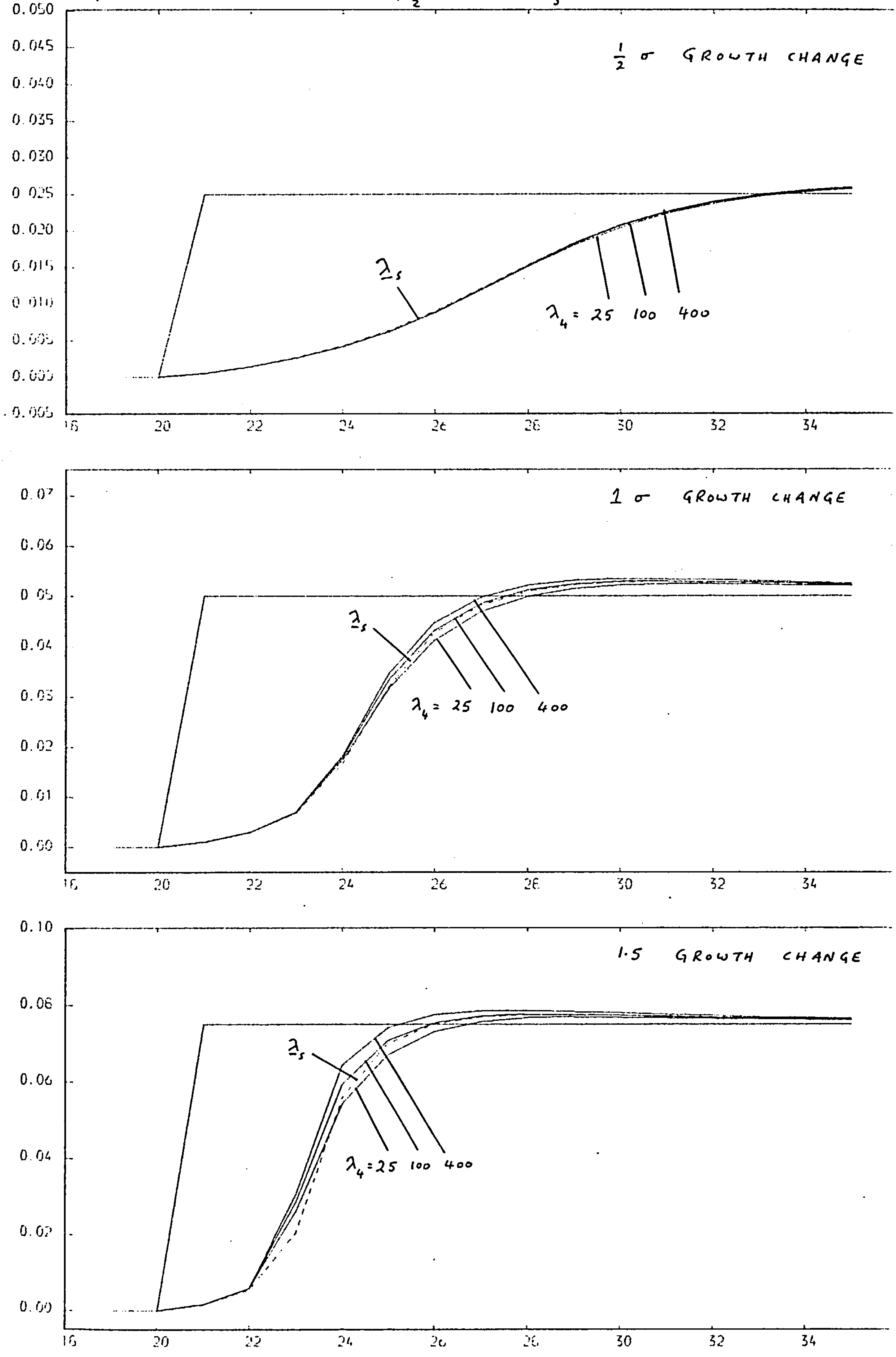
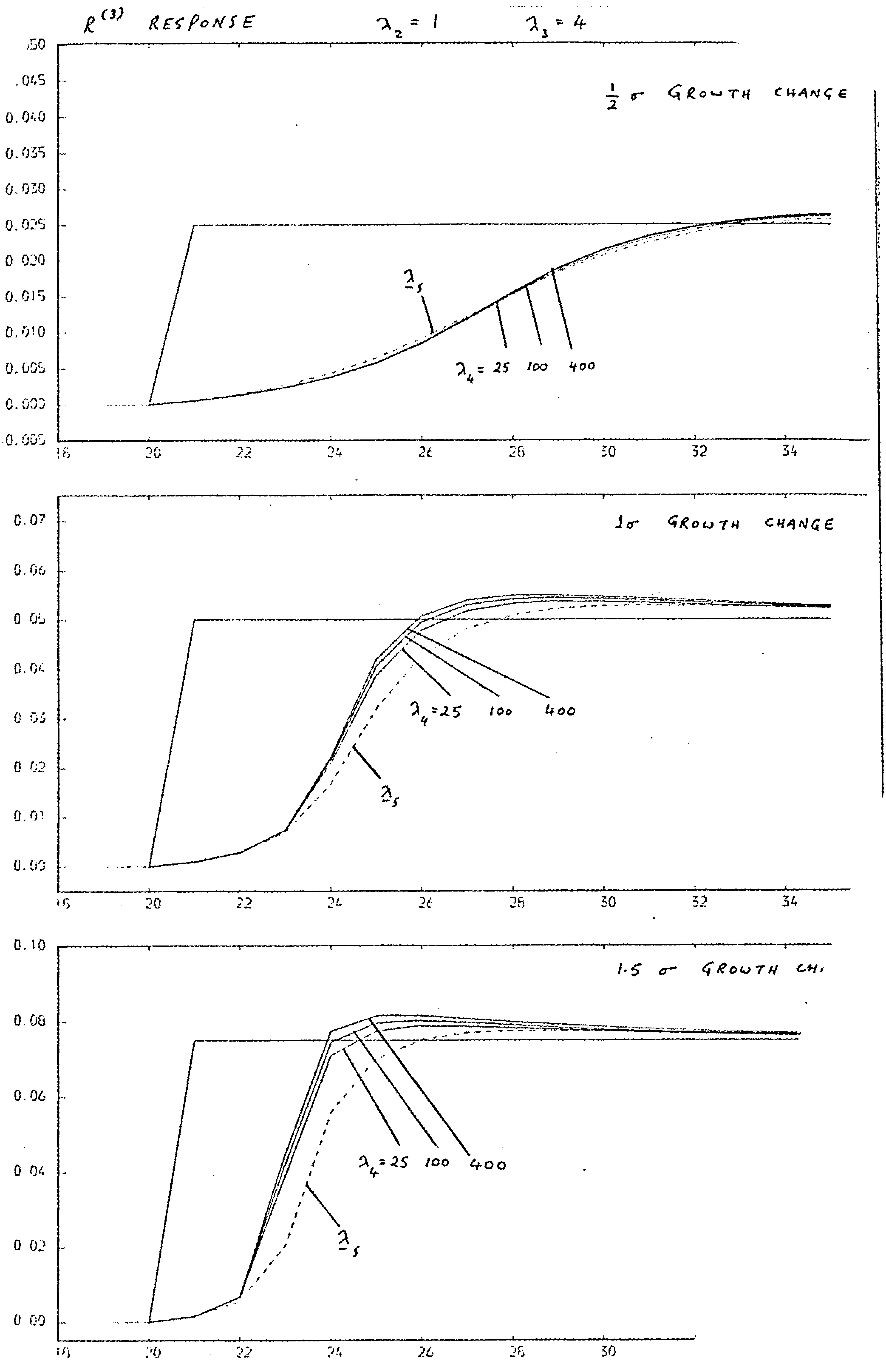


FIGURE 5.11. b



$R^{(3)}$ RESPONSE $\lambda_2 = 1$ $\lambda_3 = 4$
$\frac{1}{2}\sigma$ GROWTH CHANGE
λ_5
$\lambda_4 = 25$ 100 400
1σ GROWTH CHANGE
$\lambda_4 = 25$ 100 400
λ_5
1.5 σ GROWTH CH
$\lambda_4 = 25$ 100 400
λ_5

FIGURE 5.11.c

$R^{(4)}$ RESPONSE $\lambda_2 = 1$ $\lambda_3 = 1/4$

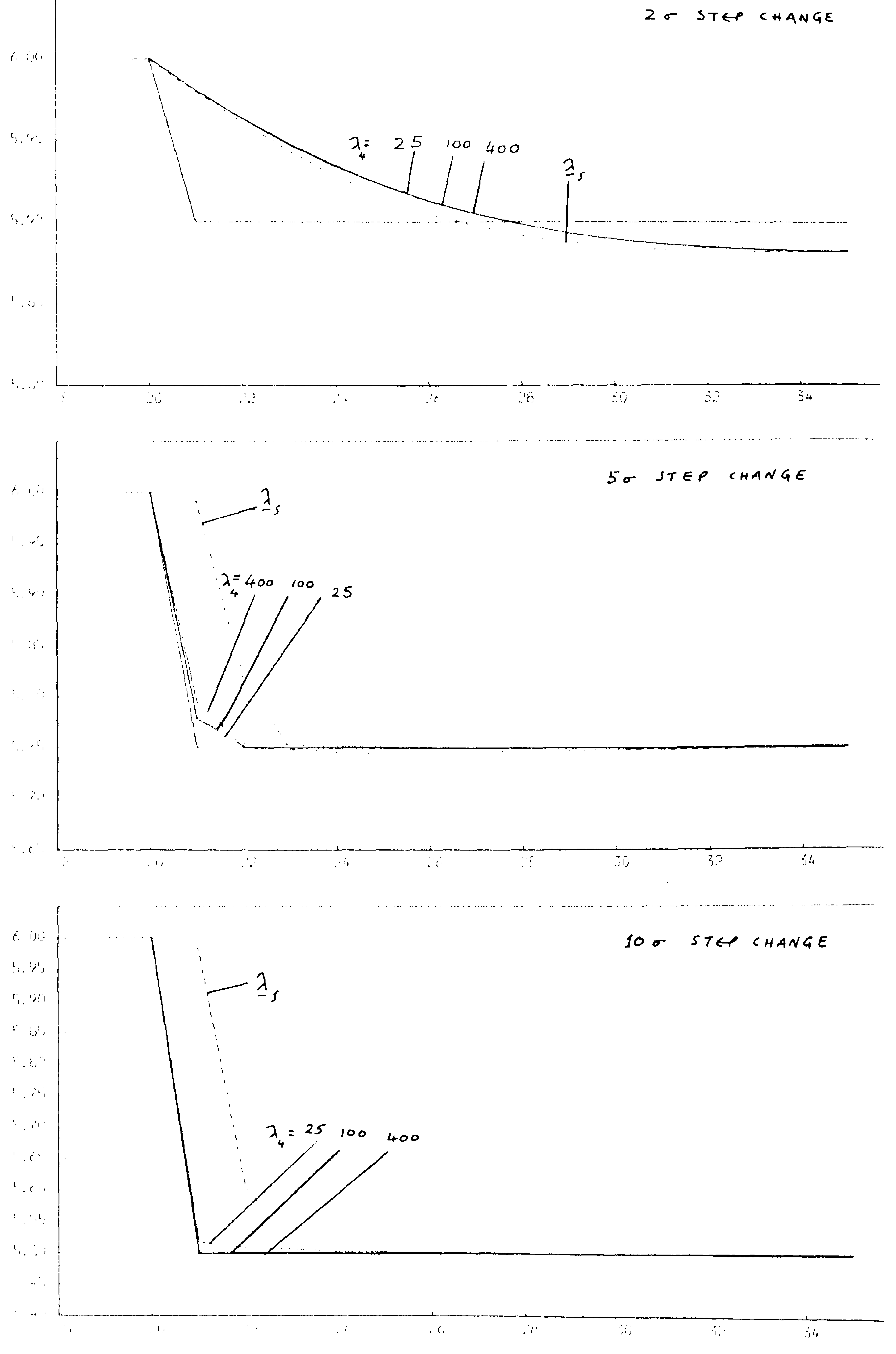


FIGURE 5.12.a

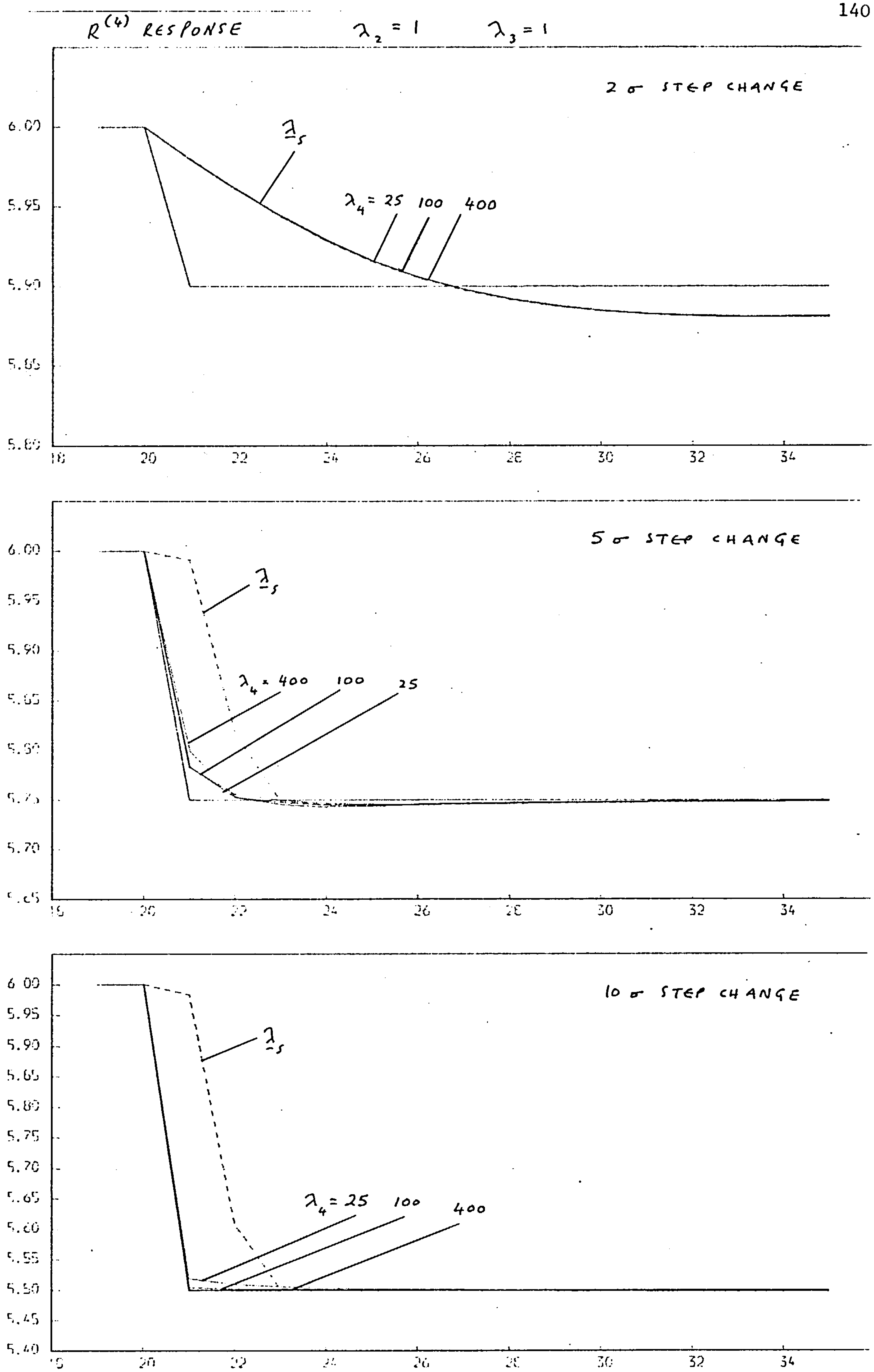
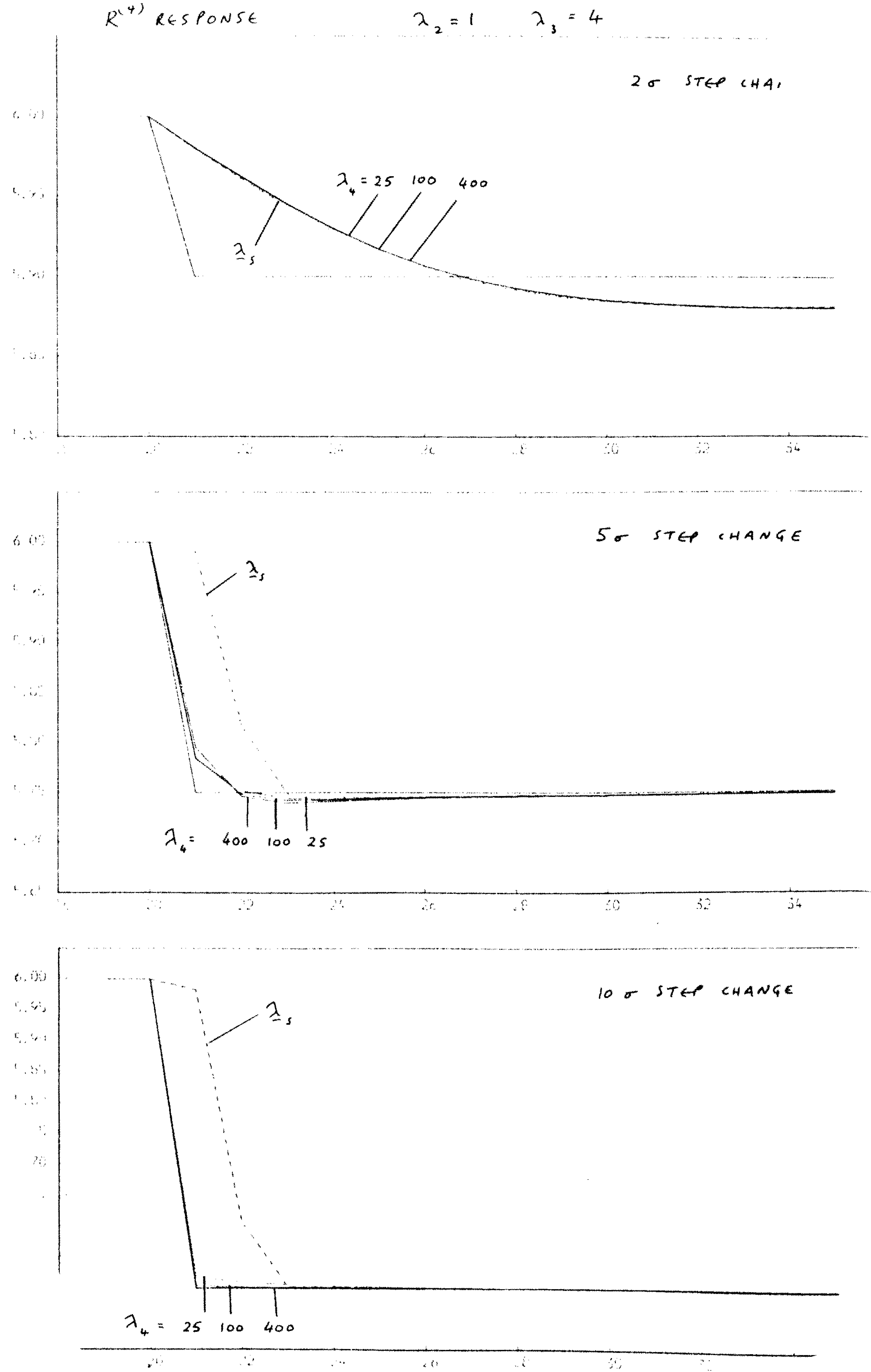

$R^{(4)}$ RESPONSE $\lambda_2 = 1$ $\lambda_3 = 1$
2 σ STEP CHANGE
λ_5
$\lambda_4 = 25$ 100 400
6.00
5.95
5.90
5.85
5.80
18
20
22
24
26
28
30
32
34
5 σ STEP CHANGE
$\lambda_4 = 400$ 100 25
10 σ STEP CHANGE
$\lambda_4 = 25$ 100 400

FIGURE 5.12.b



R^(4) RESPONSE
λ2 = 1
λ3 = 4
2σ STEP CHA
λ4 = 25
100
400
λ5
5σ STEP CHANGE
λ5
λ4 =
400
100
25
10σ STEP CHANGE
λ5
λ4 =
25
100
400

FIGURE 5.12.c

The effect of λ_3 on $R^{(4)}$ is also insignificant with λ_4 being the important factor producing a nearly identical $R^{(4)}$ to that of MSM(1,4) when λ_4 was the only factor. Figures 5.12a, 5.12b and 5.12c illustrate these points by comparison to Figure 5.8b.

5.3.7. MSM (1,2,3,4)

All four responses are now relevant and the results from a factorial experiment varying λ_2, λ_3, λ_4 are given in Table 5.16 and Figures 5.13 and 5.14. One way of analysing the responses of MSM(1,2, 3,4) is by viewing it as an extension of MSM(1,3,4) which is now required to model outliers as well as step and growth changes. We are therefore interested in the effect of λ_2 on the different responses which are to be compared with those of MSM(1,3,4).

From Table 5.16 it can be seen that small values of λ_2 (e.g. $\lambda_2 = 10$) interact with large values of λ_4 (e.g. $\lambda_4 \geq 100$) producing an undesirable z response since they interpret a large discontinuity as a step change. Values of $\lambda_2 \geq 101$ however respond well to outliers irrespective of the value of λ_4 or λ_3.

The introduction of λ_2 at a value of 101 or greater does not spoil the growth response produced by MSM(1,3,4) and some examples illustrating this are given in Figures 5.13a, 5.13b and 5.13c. These are directly comparable with 5.11a, 5.11b and 5.11c of MSM(1,3,4) respectively, the only difference being that the former have $\lambda_2 = 101$ while the latter have $\lambda_2 = 1$. The two sets of figures are almost identical suggesting that λ_2 does not affect $R^{(3)}$.

Similarly Figures 5.14a, 5.14b and 5.14c are directly comparable to 5.12a, 5.12b and 5.12c and it can be seen that the only significant difference between the two sets of $R^{(4)}$ responses is that MSM(1,2,3,4) with the introduction of $\lambda_2 = 101$ delays the system level response to a step change by one time period since it treats the first large discontinuity as an outlier rather than a step change as in MSM(1,3,4). This seems to be the only penalty in modelling outliers since as we have seen, the introduction of λ_2 does not spoil $R^{(3)}$ and additionally comparison of Tables 5.16 and 5.15 shows that $\lambda_2 = 101$ improves the MSE by approximately 2%. When λ_2 is increased considerably, to a value of 1000 for example, the response to outliers is similar to when $\lambda_2 = 101$ as can be seen from Table 5.16. The responses to growth and step changes are also very similar to those shown in Figures 5.13 and 5.14 when $\lambda_2 = 101$ and therefore are not illustrated here. However, Table 5.16 shows that the MSE when $\lambda_2 = 1000$ is marginally higher, and therefore $\lambda_2 = 101$ seems to be the best choice for MSM(1,2,3,4).

	λ_2	λ_3	λ_4	$R^{(1)}$ MSE	$R^{(2)}$ z response to outliers of size: 4σ	10σ	20σ	$R^{(3)}$, $R^{(4)}$ b_t, m_t responses
$\underline{\lambda}_s$:	101	1	100	100	.3	.3	.7	
	10	1/4	25	97	.3	.3	.2	
			100	97	2.9	4.6	3.9	
			400	97	19.3	19.8	20.0	
		1	25	99	.3	.3	.3	see Figures
			100	99	2.9	4.6	3.9	5.13a
			400	99	19.3	19.8	20.0	
		4	25	99	.4	.4	.4	5.13b
			100	99	2.9	4.6	3.9	
			400	99	19.3	19.8	20.0	5.13c
	101	1/4	25	98	.3	.2	.1	
			100	98	.2	.3	.3	and
			400	97	.1	.7	1.4	
		1	25	100	.3	.2	.1	
			100	100	.3	.3	.7	
			400	100	.2	.3	1.4	5.14a
		4	25	100	.5	.2	.1	5.14b
			100	100	.4	.3	.7	
			400	100	.4	.3	1.4	5.14c
	1000	1/4	25	99	.6	.3	.0	
			100	99	.5	.6	.3	
			400	99	.4	.4	.7	
		1	25	101	.7	.3	.0	
			100	101	.6	.6	.3	
			400	101	.5	.4	.7	
		4	25	101	1.0	.9	.8	
			100	101	.3	.6	.4	
			400	101	.0	.3	.7	

TABLE 5.16

Responses of MSM(1,2,3,4)


R(3) RESPONSE
λ2 = 101 λ3 = 1/4
1/2 σ GROWTH CHANGE
λs
λ4 = 25 100 400
1σ GROWTH CHANGE
λs
λ4 = 25 100 400
1.5 σ GROWTH CHANGE
λs
λ4 = 25 100 400

FIGURE 5.13.a

R(3) RESPONSE λ2 = 101 λ3 = 1
½σ GROWTH CHANGE
λ5
λ4 = 25 100 400
1σ GROWTH CHANGE
λ5
λ4 = 25 100 400
1.5σ GROWTH CHANGE
λ5
λ4 = 25 100 400
0.050
0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
-0.005
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0.10
0.08
0.06
0.04
0.02
0.00
18
20
22
24
26
28
30
32
34

FIGURE 5.13.6

$R^{(3)}$ RESPONSE
$\lambda_2 = 101$
$\lambda_3 = 4$
$\frac{1}{2}\sigma$ GROWTH CHANGE
λ_s
$\lambda_4 = 25$ 100 400
1σ GROWTH CHANGE
$\lambda_4 = 25$ 100 400
λ_s
$1.5\,\sigma$ GROWTH CHANGE
$\lambda_4 = 25$ 100 400
λ_s

FIGURE 5.13.c

$R^{(4)}$ RESPONSE
$\lambda_2 = 101$
$\lambda_3 = 1/4$
2σ STEP CHANGE
$\lambda_4 = 25$
100
400
$\underline{\lambda}_s$
5σ STEP CHANGE
$\lambda_4 =$
400
100
25
$\underline{\lambda}_s$
10σ STEP CHANGE
$\lambda_4 =$
25
400
100
$\underline{\lambda}_s$

FIGURE 5.14.a

$R^{(4)}$ RESPONSE $\lambda_2 = 101$ $\lambda_3 = 1$

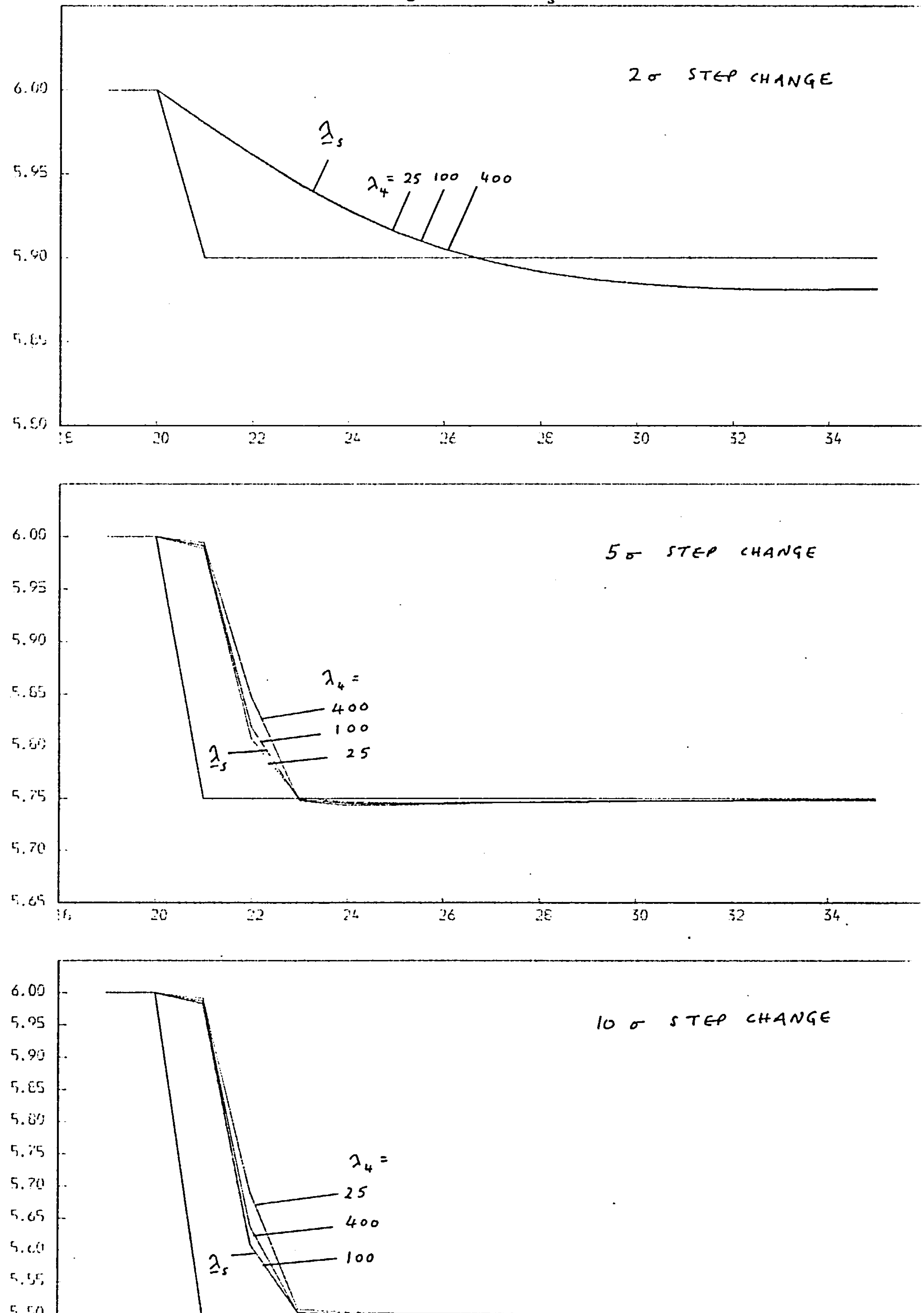


FIGURE 5.14.b


$R^{(4)}$ RESPONSE $\lambda_2 = 101$ $\lambda_3 = 4$
2 σ STEP CHANGE
λ_4 = 25 100 400
λ_s
6.00
5.95
5.90
5.85
5.80
16
20
22
24
26
28
30
32
34
5 σ STEP CHANGE
λ_4 =
400
100
25
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
10 σ STEP CHANGE
λ_4 =
25
400
100
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
5.60
5.55
5.50
5.45
5.40

FIGURE 5.14.c

5.3.8. Concluding Remarks

The standard SSP is a robust and well balanced set of parameters. Small variations from the standard SSP result in negligible change in performance while larger variations can lead to a worsening of performance across the range of discontinuities examined, which themselves cover the range of likely values from real data. No significant improvement can be made to the responses produced by MSM(1,2,3,4) with the standard SSP except in a few special cases:

(i) If no abrupt growth changes are expected then MSM(1,2,4) should be used which excludes state 3 by setting $\lambda_3 = 0$. This leads to an improvement in MSE of approximately 11% which can also be viewed as the penalty that has to be paid when using a four state model for the insurance against the possibility of the outcome of any abrupt growth changes.

(ii) If no outliers are expected then MSM(1,3,4) should be used which excludes state 2 by setting $\lambda_2 = 1$. As a result step changes are recognised almost instantaneously in contrast to the MSM(1,2,3,4) with the standard SSP which treats the first large forecast error as an outlier and consequently recognises step changes with a lag of one time period.

(iii) If neither outliers nor abrupt growth changes are expected then MSM(1,4) should be used leading to both instantaneous response to step changes and an 11% improvement in MSE.

(iv) If step changes are not expected then then MSM(1,2,3) should be used instead of MSM(1,3) even when outliers are not expected since it has smaller MSE and almost identical growth response.

The power of the four state model with the standard SSP can be illustrated by extending the comparison in Section 5.2.3 against the EWR with a discount factor $w = 0.85$. The $R^{(1)}$ and $R^{(2)}$ measures are tabulated below and the $R^{(3)}$ and $R^{(4)}$ responses graphed in Figures 5.14d and 5.14e.

	$R^{(1)}$ MSE	$R^{(2)}$ z response to outliers of size: 4σ	10σ	20σ
MSM(1,2,3,4) with $\underline{\lambda}_s$:	100	0.3	0.3	0.7
EWR with $w = 85$:	101	1.2	3.0	6.0

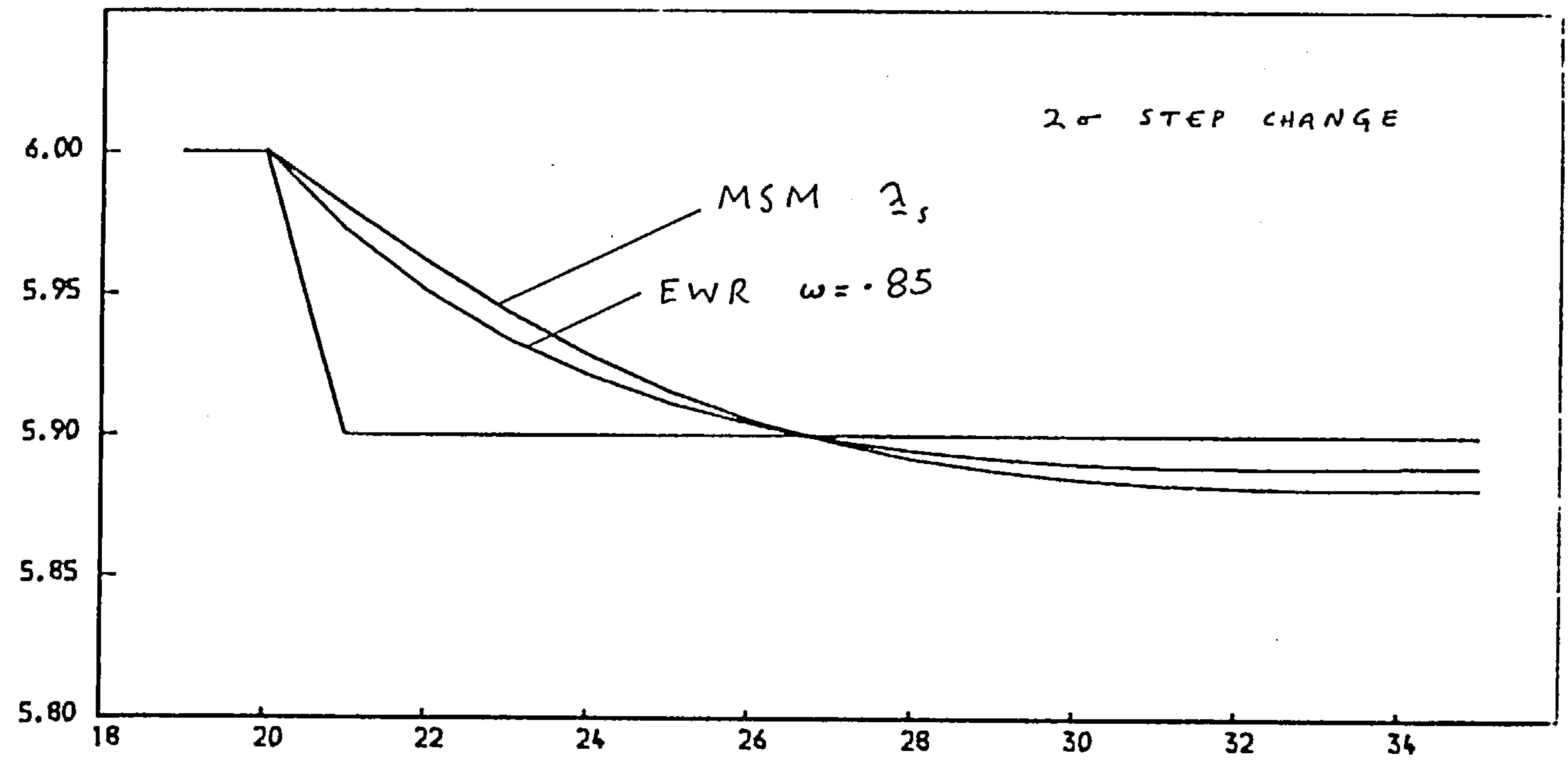
5.4. Robustness of the system to the nominated noise variance $V_{\varepsilon,N}$

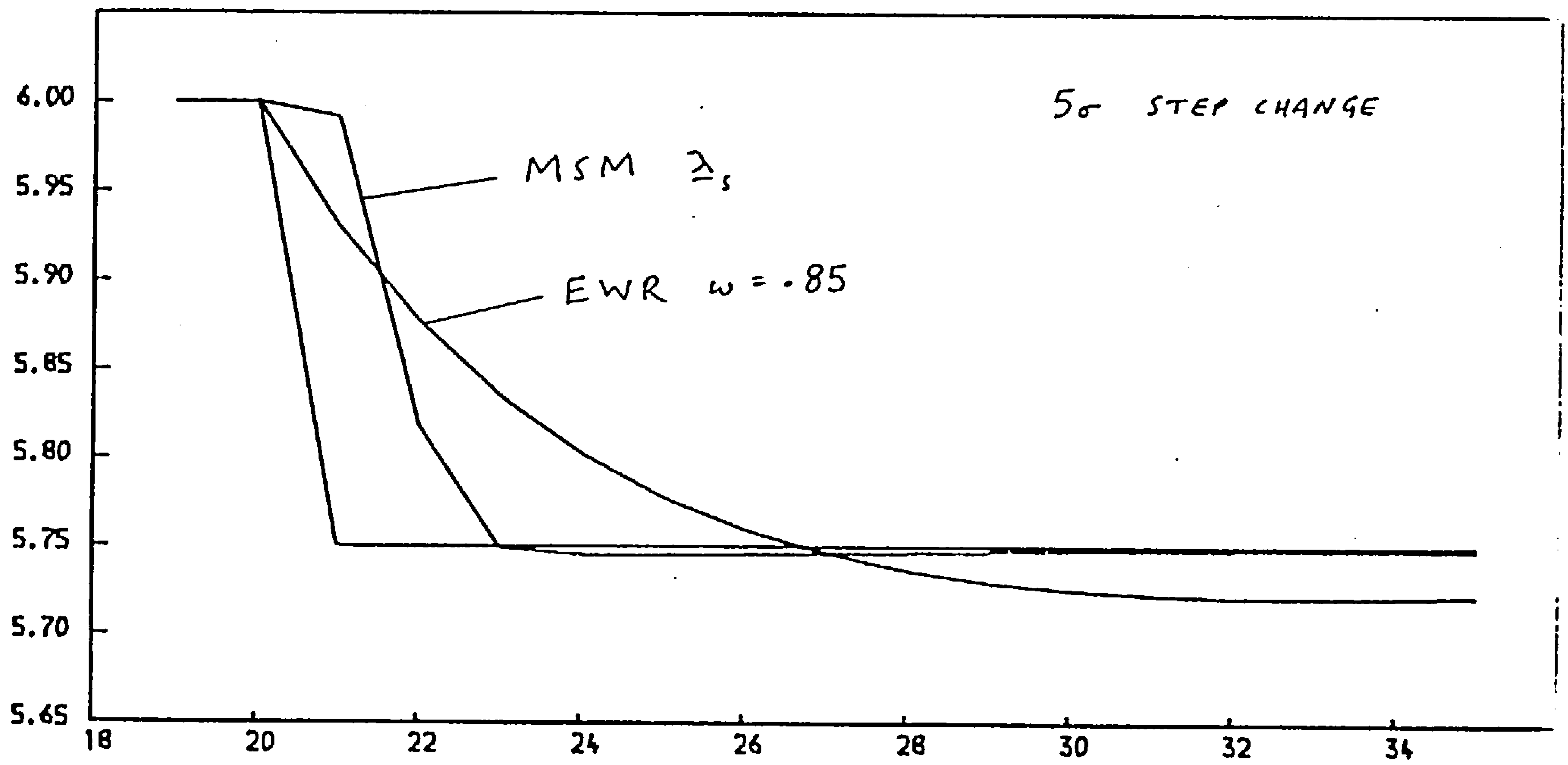
In Section 3.4.3 V_ε was defined as the true process variability during quiet periods, that is when no change is taking place. In the analysis of the system's behaviour so far V_ε was assumed to be known exactly. In practice however it is unknown and an estimate $V_{\varepsilon,N}$ must be nominated based on all available information sources including subjective judgement, experience and analogy with similar processes as well as historical data whenever available.

$R^{(3)}$ RESPONSE
$\frac{1}{2}\sigma$ GROWTH CHANGE
MSM λ_s
EWR $\omega = .85$
0.050
0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
-0.005
18
20
22
24
26
28
30
32
34
1σ GROWTH CHANGE
MSM λ_s
EWR $\omega = .85$
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
18
20
22
24
26
28
30
32
34
1.5 σ GROWTH CHANGE
MSM λ_s
EWR $\omega = .85$
0.10
0.08
0.06
0.04
0.02
0.00
18
20
22
24
26
28
30
32
34

FIGURE 5.14.d

$R^{(4)}$ RESPONSE





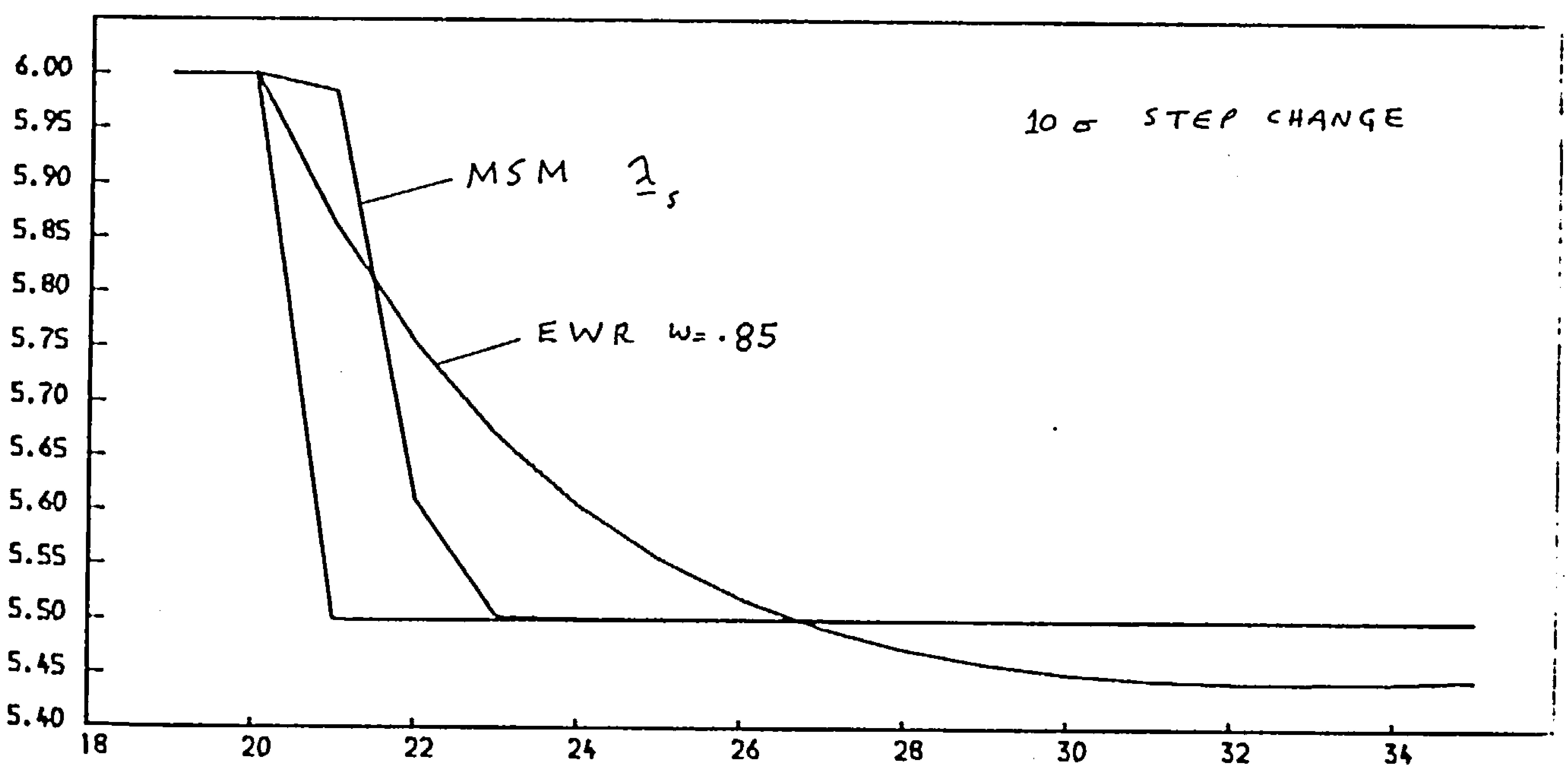


FIGURE 5.14.e

It is inevitable therefore that $V_{\varepsilon,N}$ will always be in some degree of error and the aim of this section is to test the robustness of the system's responses to errors in the nominated variance.

5.4.1. Underestimation of V_ε

It is expected that underestimation of V_ε will result in a very responsive and unstable system since during quiet periods forecast errors which are normal for the process under observation will be interpreted as unusually large. This will tend to increase uncertainty and place too much weight on recent data since these appear to the system as signalling a change in level or growth. The reason for these apparent changes is that the system interprets the data in the light of an underestimated predictive error variance.

H/S [19] report some results suggesting that the performance of the system is not significantly affected even when $V_{\varepsilon,N}$ is underestimating V_ε by a factor of four. In arriving at this conclusion they have used the standard SSP for the parameters and the Root Mean Square Error (RMSE) as the criterion of performance. Their data has been generated with $V_\varepsilon = \sigma^2 = 100$ and contains two outliers of size 10σ, one step change also of size 10σ and a growth change of size 1σ. Their results can be summarised in the following equivalent form using MSE(1) and MSE(5) to denote the mean square lead errors (cumulative demand - cumulative forecast) for lead times 1 and 5.

	MSE(1)	standardised* MSE(1)	MSE(5)	standardised* MSE(5)
$V_{\varepsilon,N} = (1/4)V_{\varepsilon,N}$	756.25	103	21258	106
$V_{\varepsilon,N} = (1/2)V_{\varepsilon}$	723.61	98.5	20164	100.6
$V_{\varepsilon,N} = V_{\varepsilon}$	734.41	100	20051	100

TABLE 5.17
Results reported by H/S
*Standardised with the value of $V_{\varepsilon,N} = V_{\varepsilon}$ case equal to 100 units

We have run a similar experiment using our "no change" data from chapter 4, which we have used so far in order to calculate $R^{(1)}$ = MSE as a measure of the system's stability during quiet periods. Our results are shown in Table 5.18 in a form directly comparable with that of Table 5.17

	MSE(1)	standardised MSE(1)	MSE(5)	standardised MSE(5)
$V_{\varepsilon,N} = (1/4)V_{\varepsilon}$	.00313	108	.0444	143
$V_{\varepsilon,N} = (1/2)V_{\varepsilon}$	.00296	102	.0354	114
$V_{\varepsilon,N} = V_{\varepsilon}$	.00292	100	.0311	100

TABLE 5.18

Comparing the two sets of results we can see from the standardised MSE(1) values that our estimates of the cost of underestimating V_ε by factors of 2 and 4 are approximately 5% higher than those of H/S. The difference is far greater when we compare the MSE with lead time 5. Underestimation by a factor of 2 now leads to a 14% penalty compared with 0.6% of H/S, while underestimation by a factor of 4 leads to a 43% penalty instead of 5%.

The results of both Tables 5.17 and 5.18 have been obtained using the standard SSP. Table 5.19 however has been obtained using the standard values for all the parameters except λ_3 which is now equal to 4 rather than its standard value of 1. This increase in λ_3 was seen in the previous section to produce a marginally faster response to growth changes without any MSE penalty. However, when V_ε is underestimated we can see by comparing Tables 5.19 and 5.18, that this increase in λ_3 leads to much higher penalties. For example, the MSE(5) penalty for $\lambda_3 = 4$ and $V_{\varepsilon,N} = (1/4)V_\varepsilon$ is 75% compared with 43% when $\lambda_3 = 1$.

		standardised MSE(1)	MSE(5)	standardised MSE(5)
$V_{\varepsilon,N} = (1/4)V_\varepsilon$	.00340	116	.0543	175
$V_{\varepsilon,N} = (1/2)V_\varepsilon$	.00305	104	.0383	124
$V_{\varepsilon,n} = V_\varepsilon$	.00293	100	.0310	100

TABLE 5.19

Our results therefore suggest that underestimation of V_ε does in fact lead to serious instability and spoils the robustness of the system to the λ's. The fact that H/S reach a different conclusion is mainly the result of using the MSE as a criterion of performance on data with discontinuities. This was seen to be misleading in Chapter 4 since a few large errors are inevitably the main contributors to the MSE value. Given that the H/S data contain 2 outliers and one step change of size 10σ each, it follows that there will be four forecast errors of size 10σ each, with a total contribution to the sum of square errors (SSE) of $4(10\sigma)^2 = 40000$, since as mentioned earlier $\sigma = 10$ for the H/S data. From Table 5.17 we can see that the SSE for lead 1 and $V_{\varepsilon,N} = V_\varepsilon$ is $60(734.41) = 44065$ since there are 60 time periods in their data. It follows therefore that the four large forecast errors account for 91% of the total SSE which to a large extent explains why the system appears robust to large errors in $V_{\varepsilon,N}$ even for five period ahead forecasts.

5.4.2. Overestimation of V_ε

Overestimation of V_ε implies that the one step ahead forecast error predictive variance will also be overestimated. As a result the system is expected to be too unresponsive since the consistently larger than normal forecast errors due to a genuine step or growth change will be interpreted as "no change" in the light of the overestimated predictive variance for the forecast error.

In figures 5.15 and 5.16 we show the system growth and system level responses for different values of $V_{\varepsilon,N}$. As expected the responses of the system when V_ε is overestimated are in some cases significantly slower than when V_ε is assumed known. From Figure 5.15 for example it can be seen that for a growth change of size 1σ the system growth when V_ε is overestimated by a factor of 4 requires on average approximately 12 time periods to cover 90% of the discontinuity compared with 7 time periods when $V_{\varepsilon,N} = V_\varepsilon$. For a growth change of size 1.5σ the equivalent comparison is 9 time periods instead of 5.

From figure 5.16 we can see that although large steps such as 10σ in size are easily recognised even when $V_{\varepsilon,N} = 4V_\varepsilon$, on smaller steps such as 5σ in size the system level responds very slowly, undershoots and remains significantly biased for over 15 time periods. The extent to which this bias is significant can be appreciated by comparing figure 5.16 with figures 5.14, 5.12 and 5.10.

Our results therefore from both figures 5.15 and 5.16 suggest that overestimation of V_ε can have serious consequences especially when medium term forecasts are required since the costs associated with the biased system level response increase with the lead time needed for forecasts. This is contrary to the conclusions suggested by the H/S results summarised in Table 5.20 below.

$R^{(3)}$ RESPONSE
STANDARD SSP
$\frac{1}{2}\sigma$ GROWTH CHANGE
$V_{\varepsilon,N} = V_\varepsilon$
$2V_\varepsilon$
$4V_\varepsilon$
1σ GROWTH CHANGE
1·5 σ GROWTH CHANGE

FIGURE 5.15

$R^{(4)}$ RESPONSE
STANDARD SSP
2 σ STEP CHANGE
$V_{\varepsilon,N} = V_\varepsilon$ $2V_\varepsilon$ $4V_\varepsilon$
5 σ STEP CHANGE
$V_{\varepsilon,N} = V_\varepsilon$ $2V_\varepsilon$ $4V_\varepsilon$
10 σ STEP CHANGE
$V_{\varepsilon,N} = V_\varepsilon$ $2V_\varepsilon$ $4V_\varepsilon$
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
5.60
5.55
5.50
5.45
5.40
20
22
24
26
28
30
32
34

FIGURE 5.16

TABLE 5.20

	MSE(1)	standardised MSE(1)	MSE(5)	standardised MSE(5)
$V_{\varepsilon,N} = V_\varepsilon$	734.41	100	20051	100
$V_{\varepsilon,N} = 2V_\varepsilon$	761.76	103.7	21170	105.6
$V_{\varepsilon,N} = 4V_\varepsilon$	789.61	107.5	23470	117

Apparently (from Table 5.20) the penalties of overestimating V_ε by factors of 2 and 4 are only 3.7% and 7.5% for the one step ahead MSE, with 5.6% and 17% the corresponding penalties for the five step ahead MSE. However, there are at least two arguments supporting the view that these penalties are grossly underestimating the true costs of overestimating V_ε. First, the size of the step change in the H/S data is large enough (10σ in size) to be easily recognised by even large overestimates of V_ε. Had the step change been of size approximately 5σ then the penalties would have been undoubtedly far more serious, as can be seen from Figure 5.16. Secondly, the fact that four large forecast errors account for such a large percentage of the MSE as was seen earlier, is again misleading. If we exclude these four errors from the calculation of MSE(1) we arrive at considerably higher penalties as shown in Table 5.21 below.

	60 forecast errors		56 forecast errors (4 10σ errors excluded)		
	MSE(1) Table 5.20	SSE(1) = 60 x MSE(1)	$SSE^*(1)$ = SSE(1)-40000	$MSE^*(1)$ = (1/56 $SSE^*(1)$)	standardised MSE*(1)
$V_{\varepsilon,N} = V_\varepsilon$	734.41	44065	4065	72.6	100
$V_{\varepsilon,N} = 2V_\varepsilon$	761.76	45706	5706	101.9	140
$V_{\varepsilon,N} = 4V_\varepsilon$	289.61	47377	7377	131.7	181

TABLE 5.21

The one step ahead MSE penalties for overestimation of V_ε by factors of 2 and 4 are now 40% and 81% respectively compared with 3.7% and 7.5% from Table 5.20.

With these comments in view, it is clear that a good estimate of V_ε is necessary in order to make a sensible statement about the true distribution of forecast errors and consequently interpret them correctly. On line estimation of V_ε is therefore essential and procedures for doing this are proposed in Chapter 6.

5.5. Concluding remarks

It has been demonstrated that in order to control any one of the four system responses it is enough to vary either the λ's or the Π's from their standard SSP values. It has also been shown that equivalences between the Π's and the λ's exist in producing a particular response and this together with the fact that some variations in the Π's can lead to large instability, suggest that the best way of controlling the system's responses is through the λ's only, with the Π's fixed at their standard SSP values. The MSM can then be viewed as a three parameter $(\lambda_2, \lambda_3, \lambda_4)$ system.

In practice a four state MSM is often required and in this case the standard SSP has been shown to be a robust and well balanced set of values for the system parameters within the range of discontinuities examined. Very large discontinuities might require larger λ's but represent such a change in the process under observation as to render time series analysis entirely inapplicable.

Except for a number of special cases, very little is gained by excluding some states (thus modelling fewer discontinuities) and conversely much is to be lost by erroneously omitting them.

Finally, a sensitivity analysis of the system's responses to errors in the nominated estimate of the true process variability showed that underestimation of V_ε can lead to serious instability. Overestimation of V_ε results in an unresponsive system taking much longer to recognise genuine changes. Furthermore overestimation of V_ε leads to serious bias in the estimates of the process level following a medium sized step change. Since in real data V_ε is unknown, methods of its on line estimation are vital for successful implementation of the MSM, and such methods are proposed in the following chapter.

CHAPTER 6

On Line Estimation of V_ε in the MSM

The observation noise variance V_ε has been shown to be critical to the performance of the MSM and on line estimation procedures are therefore essential. In this chapter two such procedures are proposed:

(i) the Fixed Range Method (FRM)

and (ii) the Variable Range Method (VRM)

Any estimation procedure implies a generating model for the variable being estimated and a simple random walk model is assumed to generate V_ε. That is, the observation noise variance at any time is assumed to have an unknown true value which is not necessarily constant through time. The changes are expected to be slow. If however there are sudden discontinuities in the variance then the methods proposed here will behave in a manner similar to how EWMA methods respond to discontinuities in the level, i.e. they are robust methods but because they do not model steps explicitly will lag behind the real value if such abrupt changes occur.

In practice major disturbances occur often from causes beyond our control such as an oil crisis, political or competitor decisions etc. In such situations or at a product launch, historical data may be quite irrelevant and on line procedures such as FRM and VRM become very important.

The basic theoretical results upon which FRM and VRM are based are described in Sections 6.1 and 6.3. These have been given in considerable detail and as a result the updating procedures look rather cumbersome. This however will completely define the system for anyone wishing to write a computer program implementing such a procedure.

Finally, numerical illustrations are given in Sections 6.2 and 6.4 using the standard SSP in order to set up the parameters of the MSM.

6.1. The Fixed Range Method (FRM)

Let us postulate that the observation noise variance at any time can be considered as having a true value which is not necessarily constant but may be slowly changing and could even have a discontinuity possibly due to external factors. We will therefore denote by $V_{\varepsilon,t}$ the true value of V_ε at time t. Hence $V_{\varepsilon,t}$ is an unobservable true parameter representing the variance of the normal distribution with zero mean from which the observation noise at time t, ε_t was obtained. We will make an estimate of $V_{\varepsilon,t}$ which will be denoted by $\hat{V}_{\varepsilon,t}$. The estimation procedure depends on a range of nominated values designated $V_\varepsilon^{(k)}$ for $k = 1,2,\ldots.K$ which will be referred to as the V range. It is then necessary to determine the probabilities $p_t^{(k)}$ that $V_\varepsilon^{(k)}$ is the correct value of the variance at time t

conditional on the observations up to time t, i.e.

$$p_t^{(k)} = p(V_\varepsilon^{(k)} | D_t) = p(V_\varepsilon^{(k)} = V_{\varepsilon,t} | D_t)$$

where $D_t = \left\{ y_1, y_2, \ldots\ldots y_t \right\}$ is the set of all observations up to and including time t. The estimate $\hat{V}_{\varepsilon,t}$ is then the probability weighted average of the values in our V range:

$$\hat{V}_{\varepsilon,t} = \sum_{k=1}^{K} p_t^{(k)} \; V_\varepsilon^{(k)}$$

Selection of values $V_\varepsilon^{(k)}$ is discussed in the next section where FRM is illustrated numerically.

We will now describe a system for updating $p_t^{(k)}$:

Using the above notation and section 3.3 we can summarise all the prior information at time t in terms of (i), (ii) and (iii) below:

$$\text{(i)} \quad \left(\begin{matrix} \mu_{t-1} \\ \beta_{t-1} \end{matrix} \,\middle|\, M_{t-1}^{(i)}, D_{t-1} \right) \sim N \left(\begin{matrix} m_{t-1}^{(i)} \\ b_{t-1}^{(i)} \end{matrix} ; \; \underline{C}_{t-1}^{(i)} \right) \qquad (6.1.1)$$

$$\text{where} \quad \underline{C}_{t-1}^{(i)} = \begin{pmatrix} c_{11,t-1}^{(i)} & c_{12,t-1}^{(i)} \\ c_{21,t-1}^{(i)} & c_{22,t-1}^{(i)} \end{pmatrix} \qquad (6.1.2)$$

and $i = 1,2,3,4$ correspond to states "no change", "outlier", "growth change" and "step change" respectively. $M_{t-1}^{(i)}$ denotes a model which defines state i at time t-1 as defined in Section 3.2. In other words (6.1.1) tells us that if $M_{t-1}^{(i)}$ is the right model at time t-1 (i.e. if the process is in state i at time t-1) then the true but unknown process parameters μ_t and β_t are best described by the normal distribution given in (6.1.1).

(ii) $\quad p_{t-1}^{(i)}$ = posterior probability of being in state i at time t (6.1.3)

and (iii) $p_{t-1}^{(k)} = p(V_{\varepsilon}^{(k)} \mid y_{t-1}, D_{t-2})$ (6.1.4)

All (i), (ii) and (iii) above are information prior to the observation at time t, y_t. From the set of equations (3.3.7) of Section 3.3 we can see that as soon as y_t becomes available there are sixteen predictive distributions corresponding to state transitions $i \to j$ from time t-1 to t. To be consistent with the notation of Section 3.3 let $M^{(ij)}$ denote a model transition from state i at time t-1 to state j at time t, i.e.

$$p(M^{(ij)} \mid D_{t-1}) = p(M_{t-1}^{(i)} \mid D_{t-1}) \cdot p(M_t^{(j)} \mid D_{t-1})$$

$$= p_{t-1}^{(i)} \cdot \Pi^{(j)} \qquad (6.1.5)$$

where $p_{t-1}^{(i)}$ is defined by (6.1.3) and $\Pi^{(j)}$ = probability of a transition to state j independent of time and process history. The latest observation y_t can then be viewed as having come from any one of these sixteen predictive distributions, i.e.

$$(y_t \mid M^{(ij)}, D_{t-1}) \sim N(\hat{y}^{(ij)}, \hat{Y}^{(ij)}) \qquad (6.1.6)$$

for $i = 1,2,3,4$ and $j = 1,2,3,4$

with $$\hat{y}^{(ij)} = m_{t-1}^{(i)} + b_{t-1}^{(i)} \qquad (6.1.7)$$

and $$\hat{Y}^{(ij)} = c_{11,t-1}^{(i)} + 2c_{12,t-1}^{(i)} + c_{22,t-1}^{(i)} + V_\varepsilon^{(j)} + V_\mu^{(j)} + V_\beta^{(j)} \qquad (6.1.8)$$

where $V_\varepsilon^{(j)}$, $V_\mu^{(j)}$ and $V_\beta^{(j)}$ can be expressed as multiples of the nominated observation noise variance $V_{\varepsilon,N}$ as shown in Table 3.1 of Section 3.4.4.

In the MSM $V_{\varepsilon,N}$ is fixed initially and remains constant through time. However, the Fixed Range Method adds another dimension to equation (6.1.8) by introducing the concept of V range consisting of values $V_\varepsilon^{(k)}$ $(k = 1,2,\ldots K)$. Hence the FRM equivalent of $\hat{Y}^{(ij)}$, is $\hat{Y}^{(ijk)}$ and uses $V_\varepsilon^{(j)}$, $V_\mu^{(j)}$, $V_\beta^{(j)}$ expressed as multiples of $V_\varepsilon^{(k)}$ (instead of $V_{\varepsilon,N}$) for $k = 1,2\ldots K$. That is, as soon as y_t becomes available it is viewed by FRM as having come from one of K sets of predictive distributions, each corresponding to a particular $V_{\varepsilon,k}$ and a particular transition $i \rightarrow j$, i.e.

$$(y_t \mid V_\varepsilon^{(k)}, M^{(ij)}, D_{t-1}) \sim N(\hat{y}^{(ij)}, \hat{Y}^{(ijk)})$$

The likelihood of y_t having been generated by one of these predictive distributions is given by:

$$L(y_t|V_\varepsilon^{(k)}, M^{(ij)}, D_{t-1}) \propto \left[\hat{Y}^{(ijk)}\right]^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}\left[e^{(ij)}\right]^2 . \left[\hat{Y}^{(ijk)}\right]^{-1}\right\} \tag{6.1.9}$$

where $e^{(ij)}$ is the forecast error at time t, i.e.,

$$e^{(ij)} = y_t - \hat{y}^{(ij)}$$

and the forecast $\hat{y}^{(ij)}$ is given by equation (6.1.7).

Our aim is now to arrive at a posterior probabilities $p_t^{(k)}$ given the latest observation y_t. Bayes' theorem can therefore be used to write:

$$p(V_\varepsilon^{(k)}|y_t, D_{t-1}) \propto p(y_t|V_\varepsilon^{(k)}, D_{t-1}) \cdot p(V_\varepsilon^{(k)}|D_{t-1})$$

and using the notation of (6.1.4),

$$p_t^{(k)} \propto p(y_t|V_\varepsilon^{(k)}, D_{t-1}) \cdot p_{t-1}^{(k)} \tag{6.1.10}$$

or $$p_t^{(k)} \propto p_{t-1}^{(k)} \sum_i \sum_j p(y_t | V_\varepsilon^{(k)}, M^{(ij)}, D_{t-1}) \cdot p(M^{(ij)} | D_{t-1}) \tag{6.1.11}$$

and therefore $p_t^{(k)}$ can now be fully determined since (i)$p_{t-1}^{(k)}$ is our prior for $V_\varepsilon^{(k)}$ given by (6.1.4), (ii) $p(y_t | V_\varepsilon^{(k)}, M^{(ij)}, D_{t-1})$ is equal to the right hand side of equation (6.1.9) times a normalising constant and (iii) $p(M^{(ij)} \mid D_{t-1})$ is given by equation (6.1.5).

Our best on line estimate of $V_{\varepsilon,t}$ can now be calculated as the weighted average of all $V_{\varepsilon}^{(k)}$:

$$\hat{V}_{\varepsilon,t} = \sum_{k} p_t^{(k)} \; V_{\varepsilon}^{(k)}$$

and this is used to update the MSM in the usual way as described in Section 3.3, the only difference being that $\hat{V}_{\varepsilon,t}$ is used instead of $V_{\varepsilon,N}$ in order to construct $V_{\varepsilon}^{(j)}$, $V_{\mu}^{(j)}$ and $V_{\beta}^{(j)}$ according to Table 3.1. Hence we can easily arrive at prior information for the next time period $t+1$, in the same form as given in equations (6.1.1) to (6.1.4) and therefore the whole procedure can be repeated with every new observation.

6.2. Illustration of the FRM

In this section we will give a numerical illustration of the FRM and its performance will be assessed and compared using the results from a number of experiments with artificial and real data. First we describe the data used for experimentation and then examine a number of different areas:

(i) Selection of values $V_{\varepsilon}^{(k)}$ for the V range as well as associated prior probabilities $p_o^{(k)}$ for $k = 1,2,\ldots K$

(ii) Choice of K

(iii) The effect of discontinuities on the on line estimates $\hat{V}_{\varepsilon,t}$ produced by the FRM. Different types and sizes of discontinuities are introduced on the same steady state realisation in order to determine the extent to which the method recognises discontinuities correctly rather than interpreting them as an increase in the observation noise variance which we are trying to estimate.

(iv) The response of the method to a discontinuity in the observation noise variance V_{ε}.

Finally the method is illustrated on two sets of real data. A listing of the computer programme used for the above experimentation is given in Appendix H.

6.2.1. Data used for experimentation.

Artificial steady state processes $\{y_t\}$ were constructed using random normal deviates generated by a NAG algorithm (see Appendix F) as follows.

Let $\{r_{t,\varepsilon}\}$, $\{r_{t,\mu}\}$ be two sequences of independent random normal deviates, i.e.

$$r_{t,\varepsilon} \sim N(0,1) \quad , \quad r_{t,\mu} \sim N(0,1)$$

Then a steady state (EWMA type) process $\{y_t\}$ can be constructed using the following generating model:

$$\left.\begin{aligned} y_t &= \mu_t + r_{t,\varepsilon}\sqrt{V_\varepsilon} \\ \mu_t &= \mu_{t-1} + r_{t,\mu}\sqrt{V_\mu} \\ &\text{for } t = 1,2,3,\ldots \end{aligned}\right\} \qquad (6.2.1.1)$$

where V_ε and V_μ are in practice the true unknown variances to be estimated. Several realisations of $\{y_t\}$ corresponding to different sequence pairs $(\{r_{t,\varepsilon}\}, \{r_{t,\mu}\})$ have been used to test the FRM but for simplicity a single typical realisation shown in Figure I.1 (see Appendix I) will be used to illustrate the method. This has been generated from (6.2.1.1) using:

$$\left.\begin{aligned} \mu_o &= 6 \\ V_\varepsilon &= .0025 \\ V_\mu &= .0001 \end{aligned}\right\} \qquad (6.2.1.2)$$

and will be referred to from now on as C6 Data (Chapter 6 Data). Where appropriate however, "average" results will be reported based on all realisations examined. Note that the variance ratio $r = V_\varepsilon/V_\mu = 25$ implies (see Section 2.3) that C6 Data can be viewed as an EWMA process with optimal $\alpha = .181$.

It should be noted here that although both estimation procedures FRM and VRM can handle a dynamic underlying true noise variance $V_{\varepsilon,t}$, our data have (for simplicity of assessing the method performance) been generated with a constant $V_{\varepsilon,t} = .0025$ and hence from now on we will refer to the true noise variance as V_ε. The advantage of a constant variance is that it allows easy comparison of the on line estimates $\hat{V}_{\varepsilon,t}$ produced by FRM and VRM, with $V_{\varepsilon,t} = V_\varepsilon = .0025$.

Discontinuities where introduced in $\{y_t\}$ at time $t = 61$ resulting in data series $\{y_t^*\}$ as follows:

(i) "outlier" of size $k_2\sigma$ at $t = 61$:

$$y_t^* = \begin{cases} y_t & t \neq 61 \\ y_t + k_2\sigma & t = 61 \end{cases}$$

(ii) "growth change" of size $k_3\sigma$ at $t = 61$:

$$y_t^* = \begin{cases} y_t & t < 60 \\ y_t + k_3\sigma(t-60) & t \geq 61 \end{cases}$$

(iii) "step change" of size $k_4\sigma$ at $t = 61$:

$$y_t^* = \begin{cases} y_t & t < 60 \\ y_t + k_4\sigma & t \geq 61 \end{cases}$$

Finally the two real data series used are as follows:

(i) Monthly total UK sales of 3M company (see C.1 in Appendix C) from 1969 to 1976. A suitable transformation has been performed for confidentiality reasons and the resulting series (seasonally adjusted) is graphed in Figure I.3 (see Appendix I). We will refer to this series from now on as 3M Data. It is interesting to note that this series contains a clear increase in variability from $t = 50$ onwards due to the uncertainty caused by 1973 world oil crisis.

(ii) Concentration readings of a chemical process reported by Box and Jenkins [3] page 525. This series is graphed in Figure I.2 (Appendix I) and will be referred to as B/J Data.

6.2.2 Selection of V range and prior probabilities

Prior to any observation at time $t = 0$ we must select values $V_\varepsilon^{(k)}$ for $k = 1,2,\ldots K$ and associated prior probabilities $p_o^{(k)}$. There is a conflict between computational effort and comprehensiveness in choosing $V_\varepsilon^{(k)}$ and $p_o^{(k)}$ but the following choice has been found to work successfully while at the same time being economically feasible:

$$\left.\begin{aligned} V_\varepsilon^{(k)} &= c^{k-6}\, V_o \quad \text{for } k = 1,2,\ldots K\ ;\quad K = 11 \\ c &= 1.5 \\ V_o &= \text{initial crude estimate of } V_\varepsilon \\ p_o^{(k)} &= \frac{1}{K} \quad \text{for } k = 1,2\ldots K \end{aligned}\right\} \qquad (6.2.2.1)$$

Hence each value $V_\varepsilon^{(k)}$ is 50% higher than the previous one and the middle value (for k = 6) is equal to V_o. An indication of the structure of $V_\varepsilon^{(k)}$ is given in Table 6.1 where the value of c^{k-6} for k = 1,2,...11 is tabulated.

k	1	2	3	4	5	6	7	8	9	10	11
c^{k-6} =	$(7.59)^{-1}$	$(5.06)^{-1}$	$(3.37)^{-1}$	$(2.25)^{-1}$	$(1.5)^{-1}$	1	1.5	2.25	3.37	5.06	7.59

TABLE 6.1

Given the V range of (6.2.2.1) it follows that the true observation noise variance V_ε will lie inside this range provided that V_o does not overestimate or underestimate it by a factor larger than 7.59. In practice errors of this magnitude are almost inconceivable and therefore this V range should be adequate almost always. However one could devise a number of decision rules which could be used to ensure that V_ε will eventually lie in our V range. It is possible for example to check at each point in time whether the latest on line estimate of V_ε, $\hat{V}_{\varepsilon,t}$, satisfies the following restriction:

$$V_\varepsilon^{(2)} < \hat{V}_{\varepsilon,t} < V_\varepsilon^{(10)} \qquad (6.2.2.2)$$

If at time $t = 20$ say, this restriction is not satisfied then we can redefine the V range using (6.2.2.1) with our best on line estimate of V_ε (i.e. $\hat{V}_{\varepsilon,20}$) in place of V_o.

An illustration of the method on C6 Data is given in Appendix J where the posterior probabilities $p_t^{(k)}$ and the on line estimation $\hat{V}_{\varepsilon,t}$ are tabulated for $t = 1,2,\ldots 200$. The posterior probabilities at time t, $p_t^{(k)}$ correspond to the following V range constructed using (6.2.2.1) with $V_o = .0060$, that is overestimating the true noise variance by a factor of just over 2.

TABLE 6.2

V range corresponding to $V_o = .0060$

k	$V_\varepsilon^{(k)}$
1	.0008
2	.0012
3	.0018
4	.0027
5	.0040
6	.0060
7	.0090
8	.0135
9	.0203
10	.0304
11	.0456

The true noise variance V_ε = .0025 lies between $V_\varepsilon^{(3)}$ and $V_\varepsilon^{(4)}$ and we therefore expect the posterior probabilities to concentrate on $p_t^{(3)}$ and $p_t^{(4)}$. This can be seen clearly in the results of Appendix J, where in the final posterior at time $t = 200$, $p_t^{(3)}$ and $p_t^{(4)}$ are equal to .34 and .66 respectively thus producing an estimate $\hat{V}_{\varepsilon,200} = .0024$ compared with the true value of $V_\varepsilon = .0025$. The $\hat{V}_{\varepsilon,t}$ estimates produced by FRM slightly underestimate V_ε since some of the genuine variability due to observation noise is to some extent wrongly interpreted as the occurrance of small discontinuities. At times the underestimation is larger and similar experiments on a large number of realisations (of the C6 Data type) have shown that the maximum factor by which $\hat{V}_{\varepsilon,t}$ can be in error from V_ε is approximately 2. However this bias is only temporary and to a large extent is the result of peculiarities in the data due to sampling error. On the whole however, the on line estimates $\hat{V}_{\varepsilon,t}$ produced by FRM are a very useful approximation to the true underlying variability.

We can now test the sensitivity of the estimates produced by FRM in relation to different errors in V_0. Consider C6 Data ($V_\varepsilon = .0025$) and four cases of V_0 being in error by factors of $\frac{1}{4}$, 1, 4 and 10 respectively, i.e.

case 1 : V_0 = .000625

case 2 : V_0 = .0025

case 3 : V_0 = .0100

case 4 : V_0 = .025

Corresponding to each case a V range can be constructed using (6.2.2.1) and these are tabulated in Table 6.3:

TABLE 6.3

k	$V_\varepsilon^{(k)}$ corresponding to case: 1	2	3	4
1	.000082	.0003	.0013	.003
2	.000123	.0005	.0020	.005
3	.000185	.0007	.0030	.007
4	.000278	.0011	.0044	.011
5	.000417	.0017	.0067	.017
6	.000625	.0025	.0100	.025
7	.000937	.0037	.0150	.037
8	.001407	.0056	.0225	.056
9	.002112	.0084	.0338	.084
10	.003170	.0127	.0506	.127
11	.004758	.0190	.0759	.190

In cases 1, 2, 3 the true noise variance $V_\varepsilon = .0025$ lies in the extreme right (between $V_\varepsilon^{(9)}$ and $V_\varepsilon^{(10)}$), centre ($V_\varepsilon^{(6)}$) and extreme left (between $V_\varepsilon^{(2)}$ and $V_\varepsilon^{(3)}$) of the V range respectively. We therefore expect the posterior probabilities $p_t^{(k)}$ to slowly concentrate around $p_t^{(9)}$, $p_t^{(6)}$, $p_t^{(2)}$. In case 4 V_ε lies outside the range to the left of $V_\varepsilon^{(1)}$ and hence we expect the posterior probabilities to concentrate on $p_t^{(1)}$.

These are illustrated in Tables 6.4 and 6.5 where $p_t^{(k)}$ and $\hat{V}_{\varepsilon,t}$ corresponding to cases 1,2,3,4 are tabulated for times $t = 1,5,10,15,\ldots 100$. It can be seen that as time progresses $p_t^{(k)}$ concentrate on the expected places in each of the V ranges corresponding to cases 1,2,3 and 4. It is interesting to note that after a small

Table 6.4

t	$\rho_t^{(1)}$	$\rho_t^{(2)}$	$\rho_t^{(3)}$	$\rho_t^{(4)}$	$\rho_t^{(5)}$	$\rho_t^{(6)}$	$\rho_t^{(7)}$	$\rho_t^{(8)}$	$\rho_t^{(9)}$	$\rho_t^{(10)}$	$\rho_t^{(11)}$	$\hat{V}_{\varepsilon,t}$
1	.10	.09	.09	.09	.09	.09	.09	.09	.09	.09	.08	.0012
5	.11	.11	.12	.12	.12	.11	.10	.08	.06	.04	.02	.0008
10	.02	.02	.03	.03	.05	.09	.15	.21	.20	.14	.06	.0017
15				.01	.01	.03	.08	.20	.31	.25	.11	.0023
20						.01	.05	.19	.36	.28	.09	.0024
25					.01	.02	.08	.27	.38	.20	.04	.0021
30					.01	.03	.11	.33	.37	.13	.01	.0019
35					.01	.04	.16	.40	.32	.06		.0016
40					.01	.04	.14	.29	.33	.17	.02	.0019
45					.01	.04	.16	.33	.32	.13	.01	.0018
50						.03	.13	.34	.37	.12	.01	.0018
55						.01	.07	.31	.45	.15	.01	.0020
60							.03	.27	.51	.18	.01	.0021
65							.04	.32	.51	.12		.0020
70							.01	.15	.58	.25	.01	.0023
75							.02	.25	.59	.14		.0021
80							.02	.30	.58	.09		.0020
85							.01	.22	.65	.13		.0021
90							.01	.20	.67	.12		.0021
95								.16	.71	.12		.0021
100							.01	.25	.68	.07		.0020

case 1 : $V_o = \cdot 000625$

t	$\rho_t^{(1)}$	$\rho_t^{(2)}$	$\rho_t^{(3)}$	$\rho_t^{(4)}$	$\rho_t^{(5)}$	$\rho_t^{(6)}$	$\rho_t^{(7)}$	$\rho_t^{(8)}$	$\rho_t^{(9)}$	$\rho_t^{(10)}$	$\rho_t^{(11)}$	$\hat{V}_{\varepsilon,t}$
1	.10	.10	.10	.10	.09	.09	.09	.09	.09	.08	.07	.0046
5	.20	.18	.16	.14	.11	.08	.06	.04	.02	.01	.01	.0016
10	.09	.11	.15	.18	.18	.15	.09	.04	.01			.0018
15	.02	.03	.06	.13	.24	.27	.18	.07	.02			.0025
20		.01	.03	.10	.25	.34	.20	.06	.01			.0025
25	.01	.02	.05	.16	.33	.31	.12	.02				.0021
30	.01	.02	.07	.21	.38	.26	.06					.0018
35	.01	.03	.09	.28	.39	.18	.02					.0016
40	.01	.03	.09	.22	.30	.25	.09	.01				.0018
45	.01	.03	.10	.24	.32	.23	.06					.0017
50		.02	.07	.23	.37	.26	.05					.0018
55			.03	.16	.41	.34	.06					.0020
60			.01	.10	.41	.41	.07					.0021
65			.01	.13	.46	.36	.04					.0020
70				.03	.33	.55	.09					.0023
75				.07	.45	.44	.04					.0021
80				.09	.51	.38	.02					.0020
85				.05	.46	.47	.03					.0021
90				.04	.46	.48	.02					.0021
95				.02	.43	.53	.02					.0021
100				.04	.55	.40	.01					.0020

case 2 : $V_o = \cdot 0025$

Table 6.5

t	$\rho_t^{(1)}$	$\rho_t^{(2)}$	$\rho_t^{(3)}$	$\rho_t^{(4)}$	$\rho_t^{(5)}$	$\rho_t^{(6)}$	$\rho_t^{(7)}$	$\rho_t^{(8)}$	$\rho_t^{(9)}$	$\rho_t^{(10)}$	$\rho_t^{(11)}$	$\hat{V}_{\varepsilon,t}$
1	.11	.11	.11	.11	.10	.10	.09	.08	.07	.06	.06	.0160
5	.26	.22	.18	.13	.09	.06	.03	.02	.01			.0044
10	.35	.28	.19	.11	.05	.02						.0027
15	.25	.30	.25	.14	.05	.01						.0027
20	.20	.32	.30	.14	.04							.0027
25	.31	.38	.24	.07	.01							.0022
30	.41	.40	.17	.03								.0020
35	.52	.37	.10	.01								.0017
40	.37	.35	.22	.05								.0021
45	.43	.36	.18	.03								.0019
50	.42	.39	.17	.02								.0019
55	.36	.43	.19	.02								.0020
60	.27	.48	.23	.02								.0021
65	.33	.49	.17	.01								.0019
70	.13	.51	.34	.02								.0023
75	.23	.56	.20	.01								.0020
80	.28	.57	.15									.0019
85	.19	.61	.20									.0021
90	.17	.63	.19									.0021
95	.13	.66	.21									.0021
100	.21	.67	.12									.0020

case 3 : $V_0 = .0100$

t	$\rho_t^{(1)}$	$\rho_t^{(2)}$	$\rho_t^{(3)}$	$\rho_t^{(4)}$	$\rho_t^{(5)}$	$\rho_t^{(6)}$	$\rho_t^{(7)}$	$\rho_t^{(8)}$	$\rho_t^{(9)}$	$\rho_t^{(10)}$	$\rho_t^{(11)}$	$\hat{V}_{\varepsilon,t}$
1	.13	.12	.12	.11	.10	.10	.08	.07	.06	.05	.04	.0350
5	.32	.25	.18	.12	.07	.04	.02	.01				.0087
10	.52	.29	.13	.05	.01							.0050
15	.59	.29	.10	.02								.0044
20	.65	.27	.06	.01								.0041
25	.79	.19	.02									.0037
30	.87	.12	.01									.0035
35	.93	.07										.0034
40	.87	.12	.01									.0035
45	.91	.09										.0034
50	.94	.06										.0034

case 4 : $V_0 = .0250$

initial period the on line estimates $\hat{V}_{\varepsilon,t}$ are almost identical in each of the first three cases despite the fact that the posterior probabilities are all quite different as a result of the different V ranges. This illustrates the fact that the position of V_ε in our V range is not critical or alternatively that the effect of our initial crude estimate V_o dies away very quickly and $\hat{V}_{\varepsilon,t}$ is then wholly determined by the actual data.

In case 4 V_ε lies outside the V range and as a result the probabilities very quickly concentrate on the extreme left of the range. If the decision rule described earlier was used in this case, then it can be seen (Table 6.4) that at $t = 15$ the restriction given by (6.2.2.2) would not be satisfied since $V_{\varepsilon,15} < V_\varepsilon^{(2)}$. A new V range could therefore be constructed using $V_o = \hat{V}_{\varepsilon,15} = .0044$ which implies that V_ε would lie very near the middle of this new range and if the FRM procedure were applied again from $t=1$ the results would be very similar to those of cases 1, 2 and 3 earlier.

6.2.3. Effect of the choice of K

The value of $K = 11$ has so far been used and in this section we examine the effect of larger K thus increasing the number of $V_\varepsilon^{(k)}$ values in the V range. Consider the following three cases for K and c:

case 1 : $K = 11$ $c = 1.5$

case 2 : $K = 23$ $c = 1.21$

case 3 : $K = 45$ $c = 1.1$

where the c's have been chosen such that in each case the extreme values of the V range, $V_\varepsilon^{(1)}$ and $V_\varepsilon^{(K)}$, are approximately $\frac{1}{8}V_o$ and $8V_o$ respectively. $V_\varepsilon^{(k)}$ and prior probabilities $p_o^{(k)}$ are then calculated from,

$$\left.\begin{aligned} V_\varepsilon^{(k)} &= V_o c^{(k-\frac{K+1}{2})} \\ p_o^{(k)} &= \frac{1}{K} \\ \text{for } k &= 1,2,\ldots K \end{aligned}\right\} \qquad (6.2.3.1)$$

Note that for $K = 11$ (6.2.3.1) reduces to (6.2.2.1). The performance of FRM for the three cases (all of which assume $V_o = 4V_\varepsilon = .0100$) is illustrated in Table 6.6 showing $\hat{V}_{\varepsilon,t}$ for $t = 1,5,10,\ldots 100$ and Table 6.7 showing $V_\varepsilon^{(k)}$ together with the associated probabilities $p_t^{(k)}$ at time $t = 100$ and for $k = 1,2,\ldots K$.

	case 1	case 2	case 3
	K=11	K=23	K=45
t	$\hat{V}_{\varepsilon,t}$	$\hat{V}_{\varepsilon,t}$	$\hat{V}_{\varepsilon,t}$
1	.0160	.0158	.0156
5	.0044	.0045	.0046
10	.0027	.0027	.0028
15	.0027	.0028	.0028
20	.0027	.0027	.0027
25	.0022	.0022	.0023
30	.0020	.0020	.0020
35	.0017	.0018	.0018
40	.0021	.0021	.0021
45	.0019	.0020	.0020
50	.0019	.0019	.0020
55	.0020	.0020	.0020
60	.0021	.0021	.0021
65	.0019	.0020	.0020
70	.0023	.0023	.0023
75	.0020	.0020	.0021
80	.0019	.0019	.0020
85	.0021	.0021	.0021
90	.0021	.0021	.0021
95	.0021	.0021	.0021
100	.0020	.0020	.0020

Table 6.6

Table 6.7

k	$V_\varepsilon^{(k)}$	$P_{100}^{(k)}$	$V_\varepsilon^{(k)}$	$P_{100}^{(k)}$	$V_\varepsilon^{(k)}$	$P_{100}^{(k)}$
1	.0013	.2101	.0012	.0664	.0012	.0346
2	.0020	.6700	.0015	.1847	.0014	.0608
3	.0030	.1193	.0018	.3062	.0015	.0947
4	.0044	.0006	.0022	.2808	.0016	.1298
5	.0067	.0000	.0026	.1307	.0018	.1551
6	.0100	.0000	.0032	.0284	.0020	.1600
7	.0150	.0000	.0039	.0027	.0022	.1410
8	.0225	.0000	.0047	.0001	.0024	.1049
9	.0338	.0000	.0056	.0000	.0026	.0651
10	.0506	.0000	.0068	.0000	.0029	.0334
11	.0759	.0000	.0083	.0000	.0032	.0141
12			.0100	.0000	.0035	.0048
13			.0121	.0000	.0039	.0014
14			.0146	.0000	.0042	.0003
15			.0177	.0000	.0047	.0001
16			.0214	.0000	.0051	.0000
17			.0259	.0000	.0056	.0000
18			.0314	.0000	.0062	.0000
19			.0380	.0000	.0068	.0000
20			.0459	.0000	.0075	.0000
21			.0556	.0000	.0083	.0000
22			.0673	.0000	.0091	.0000
23			.0814	.0000	.0100	.0000
24					.0110	.0000
25					.0121	.0000
26					.0133	.0000
27					.0146	.0000
28					.0161	.0000
29					.0177	.0000
30					.0195	.0000
31					.0214	.0000
32					.0236	.0000
33					.0259	.0000
34					.0285	.0000
35					.0314	.0000
36					.0345	.0000
37					.0380	.0000
38					.0418	.0000
39					.0459	.0000
40					.0505	.0000
41					.0556	.0000
42					.0612	.0000
43					.0673	.0000
44					.0740	.0000
45					.0814	.0000
	case 1, $K=11$		case 2, $K=23$		case 3, $K=45$	

The one line estimates $\hat{V}_{\varepsilon,t}$ are almost identical and at time $t = 100$ $\hat{V}_{\varepsilon,100}$ equals .0020 in all cases even though the posterior probabilities and V ranges are completely different. We can therefore conclude that the choice of c and K is not critical to the performance of FRM provided $K \geq 11$ and $c \leq 1.5$ and the particular case of $k = 11$ and $c = 1.5$ can be recommanded on grounds of minimum computational effort.

6.2.4. Effect of process discontinuities

The point of interest here is to assess the extent to which the on line estimates $\hat{V}_{\varepsilon,t}$ are affected by the occurrence of abrupt changes in the underlying true process.

Let us denote by $\hat{V}_{\varepsilon,t,*}$ the on line V_ε estimate produced by FRM at time t, when applied to C6 Data (or any other series without abrupt changes). Hence $\hat{V}_{\varepsilon,t,*}$ has the same meaning as $\hat{V}_{\varepsilon,t}$ except that it implies that the data used have no discontinuities. We can then assess the extent to which $\hat{V}_{\varepsilon,t}$ is affected by discontinuities, by comparing $\hat{V}_{\varepsilon,t,*}$ against $\hat{V}_{\varepsilon,t}$ produced by FRM and corresponding to a particular type and size of discontinuity.

Different types and sizes of discontinuities have been introduced in C6 Data at time $t = 61$ as described earlier in 6.2.2, and in Table 6.8 we have tabulated $\hat{V}_{\varepsilon,t,*}$ and $\hat{V}_{\varepsilon,t}$ for $t = 50,51,\ldots100$, using $V_o = .0050$, that is initially overestimating the true noise variance by 100%. Note that before $t = 61$ all $\hat{V}_{\varepsilon,t}$ are the same

Table 6.8

t	$\hat{V}_{\varepsilon,t,*} \times 10^4$	$\hat{V}_{\varepsilon,t} \times 10^4$: OUTLIER of size: 4σ	10σ	20σ	GROWTH CHANGE of size: $\frac{1}{2}\sigma$	1σ	$1{\cdot}5\sigma$	STEP CHANGE of size: 2σ	5σ	10σ
51	17	17	17	17	17	17	17	17	17	17
52	18	18	18	18	18	18	18	18	18	18
53	20	20	20	20	20	20	20	20	20	20
54	20	20	20	20	20	20	20	20	20	20
55	19	19	19	19	19	19	19	19	19	19
56	20	20	20	20	20	20	20	20	20	20
57	19	19	19	19	19	19	19	19	19	19
58	21	21	21	21	21	21	21	21	21	21
59	21	21	21	21	21	21	21	21	21	21
60	21	21	21	21	21	21	21	21	21	21
61	20	20	21	25	20	21	22	22	20	21
62	19	20	20	24	20	22	25	23	20	21
63	19	19	20	24	19	22	26	25	19	21
64	20	20	21	24	22	22	26	24	20	21
65	19	20	20	24	23	22	26	24	20	21
66	19	19	20	23	24	22	26	23	19	20
67	20	21	21	24	26	22	26	24	19	21
68	20	21	21	24	26	23	26	24	20	21
69	21	21	22	25	26	23	26	25	20	22
70	23	23	24	26	27	25	27	26	22	24
71	22	23	23	26	27	24	27	25	22	23
72	22	22	23	25	27	24	27	25	21	23
73	22	22	22	25	26	23	26	25	21	22
74	21	21	22	24	26	23	25	25	20	22
75	21	21	21	24	25	22	25	24	20	21
76	20	20	21	23	24	21	24	23	19	21
77	20	20	21	23	24	21	24	24	19	21
78	20	20	21	23	24	21	24	24	19	21
79	20	20	20	23	24	21	24	23	19	20
80	20	20	20	23	24	20	23	23	19	20
81	21	22	22	24	25	23	25	25	21	22
82	21	21	22	24	25	23	25	25	21	22
83	21	21	22	24	24	22	24	24	20	22
84	20	20	21	23	24	21	24	24	20	21
85	21	21	21	23	24	22	24	24	20	21
86	21	21	21	23	24	22	24	24	20	21
87	21	21	21	23	24	22	24	24	20	21
88	20	21	21	23	24	21	24	24	20	21
89	21	21	22	24	25	22	24	24	21	22
90	21	21	21	23	24	22	24	24	20	21
91	20	21	21	23	24	21	23	23	20	21
92	20	20	21	23	23	21	23	23	20	21
93	20	21	21	23	23	21	23	23	20	21
94	21	22	22	24	24	22	24	24	21	22
95	21	21	22	23	24	22	24	24	21	22
96	21	21	22	23	24	22	24	24	21	22
97	21	21	21	23	24	21	23	23	20	21
98	20	20	21	23	23	21	23	23	20	21
99	20	20	21	22	23	21	23	23	19	21
100	20	20	20	22	22	20	22	22	19	20

$\hat{V}_{\varepsilon,t} \times 10^4$ produced by FRM when applied to C6 Data with discontinuities at $t=61$

$\hat{V}_{\varepsilon,t,*} \times 10^4$ has been produced by FRM when applied to C6 Data without any discontinuity at $t=61$

since the discontinuity is at $t = 61$. It can be seen that the effect of discontinuities considered, results in a small overestimation of $\hat{V}_{\varepsilon,t,*}$, the largest difference $(\hat{V}_{\varepsilon,t} - \hat{V}_{\varepsilon,t,*})$ being of the order of 20%. This is because at points of abrupt change the system receives a number of unusually large forecast errors and consequently part of the uncertainty is interpreted by the FRM as an increase in noise variance. As the system adapts to the new situation, $\hat{V}_{\varepsilon,t}$ slowly converges back to $\hat{V}_{\varepsilon,t,*}$ levels. We can therefore conclude that FRM is a robust variance estimation procedure and is not significantly affected by the type and size of discontinuities most frequently observed in the real world.

6.2.5. Effect of a discontinuity in V_ε

We now investigate the response of FRM when there is a discontinuity in the noise variance. In Section 6.2.1 we described how C6 Data were generated using $V_\varepsilon = .0025$ for $t = 1,2,\ldots 200$. Consider now C6 Data again, this time generated using

$$\left.\begin{array}{llll} V_\varepsilon & = & .0025 & t \leq 100 \\ V_\varepsilon & = & .0050 & t > 100 \end{array}\right\} \qquad (6.2.5.1)$$

Consider also the following V range constructed using (6.2.2.1) with $V_o = .0050$:

TABLE 6.9

k	V range corresponding to $V_o = .0050$ $V_\varepsilon^{(k)}$
1	.0007
2	.0010
3	.0015
4	.0022
5	.0033
6	.0050
7	.0075
8	.0112
9	.0169
10	.0253
11	.0380

In Table 6.10 we give the posterior probabilities $p_t^{(k)}$ and on line estimates $\hat{V}_{\varepsilon,t}$ produced by FRM using the V range of Table 6.9 and C6 Data generated by V_ε as given in (6.2.5.1) i.e. with a discontinuity at time $t = 101$. Table 6.10 shows that at time $t = 100$ the posterior probabilities are concentrated on $p_t^{(3)} = .35$ and $p_t^{(4)} = .60$ corresponding to $V_\varepsilon^{(3)} = .0015$ and $V_\varepsilon^{(4)} = .0022$ respectively and leading to $\hat{V}_{\varepsilon,100} = .0020$. From $t = 101$ onwards however, the discontinuity in noise variance means that its new level of $V_\varepsilon = .0025$ lies on $V_\varepsilon^{(6)} = .0050$. After $t = 100$ $V_\varepsilon^{(5)}$ and $V_\varepsilon^{(6)}$ will therefore have higher likelihoods than $V_\varepsilon^{(3)}$ and $V_\varepsilon^{(4)}$ and hence we expect the posterior probabilities to shift towards $p_t^{(5)}$, $p_t^{(6)}$ and away from $p_t^{(3)}$, $p_t^{(4)}$. This shifting of probabilities is clearly shown in the results of Table 6.10 after $t = 100$.

t	$\rho_t^{(1)}$	$\rho_t^{(2)}$	$\rho_t^{(3)}$	$\rho_t^{(4)}$	$\rho_t^{(5)}$	$\rho_t^{(6)}$	$\rho_t^{(7)}$	$\rho_t^{(8)}$	$\rho_t^{(9)}$	$\rho_t^{(10)}$	$\rho_t^{(11)}$	$\hat{V}_{\varepsilon,t}$
95		.01	.24	.67	.08							
100		.02	.35	.60	.04							.0021
105		.01	.19	.69	.12							.0020
110			.16	.71	.12							.0022
115			.16	.73	.11							.0022
120			.06	.62	.31							.0022
125			.02	.44	.52	.01						.0025
130			.01	.35	.62	.02						.0028
135			.01	.33	.65	.02						.0029
140			.01	.36	.61	.01						.0030
145			.01	.28	.70	.01						.0029
150			.01	.29	.70	.01						.0030
155				.27	.72	.01						.0030
160				.19	.79	.02						.0030
165				.13	.85	.02						.0031
170				.14	.84	.02						.0032
175				.16	.82	.01						.0032
180				.09	.89	.02						.0032
185				.11	.88	.01						.0033
190				.12	.87	.02						.0032
195				.04	.92	.04						.0032
200				.03	.92	.05						.0034
												.0034

Table 6.10

t	$\rho_t^{(1)}$	$\rho_t^{(2)}$	$\rho_t^{(3)}$	$\rho_t^{(4)}$	$\rho_t^{(5)}$	$\rho_t^{(6)}$	$\rho_t^{(7)}$	$\rho_t^{(8)}$	$\rho_t^{(9)}$	$\rho_t^{(10)}$	$\rho_t^{(11)}$	$\hat{V}_{\varepsilon,t}$
95		.01	.24	.66	.09							
100		.02	.35	.58	.04							.0022
105		.01	.18	.67	.13	.01						.0020
110		.01	.16	.69	.13	.01						.0024
115		.01	.16	.71	.12							.0023
120			.05	.51	.30	.04	.04	.03	.02	.01	.01	.0023
125			.01	.23	.28	.08	.14	.13	.08	.04	.01	.0038
130			.01	.14	.24	.09	.21	.19	.09	.03	.01	.0071
135			.01	.21	.34	.11	.18	.11	.03			.0077
140			.02	.33	.42	.09	.10	.03				.0054
145			.01	.28	.47	.11	.10	.02				.0038
150			.01	.31	.51	.10	.06	.01				.0038
155			.01	.31	.54	.10	.04					.0035
160				.21	.59	.13	.05					.0033
165				.15	.61	.17	.06					.0036
170				.18	.65	.13	.03					.0038
175				.21	.67	.10	.01					.0035
180				.12	.70	.15	.03					.0033
185				.15	.72	.11	.01					.0036
190				.15	.69	.13	.02					.0034
195				.04	.57	.28	.08	.01	.01	.01		.0035
200				.03	.57	.30	.08	.01				.0046
												.0043

Table 6.11 $(p_L = .0001)$

FRM applied to C6 Data with $V_t = \begin{cases} .0025 & t \le 100 \\ .0050 & t > 100 \end{cases}$

At $t = 150$ for example $p_t^{(3)}$ has gone down to a value of .01 (from .35 at $t = 100$) while $p_t^{(5)}$ has increased to .70 (from .04 at $t = 100$).

However this process of $p_t^{(k)}$ redistribution can sometimes be quite slow because of the way posterior probabilities are calculated. From equation (6.1.10) we can write the posterior $p_t^{(k)}$ as follows:

$$p_t^{(k)} \propto L_t^{(k)} \cdot p_{t-1}^{(k)} \qquad (6.2.5.2.)$$

where $L_t^{(k)}$ is the likelihood of the observation at time t conditional on $V_\varepsilon^{(k)}$ and model transition i to j (see (6.1.9)). From Table 6.10 however we can see that after $t = 100$ when the method should respond to the new V_ε level, $p_{t-1}^{(k)}$ for $k > 5$ are so small that no matter how large $L_t^{(k)}$ may be, it makes a small contribution to $p_t^{(k)}$ and hence the speed of response may be unsatisfactory.

In order to improve the speed of response we propose the introduction of a lower limit p_L on the posteriors $p_t^{(k)}$ so that $p_t^{(k)} \geq p_L$ at all times. Clearly, the results reported so far correspond to $p_L = 0$. We now seek a level of p_L which will be small enough not to affect $\hat{V}_{\varepsilon,t}$ during periods of relative stability in V_ε but help the updating procedure to shift the probabilities faster when a genuine and abrupt V_ε change has taken place.

Consider the posterior probabilities $p_{t-1}^{(k)}$ calculated from equation (6.1.11) at time $t-1$. Then we use the following transformation in order to produce slightly different posteriors:

$$\left.\begin{aligned} p_{t-1}^{(k)*} &= p_L + (1 - Kp_L)\ p_{t-1}^{(k)} \\ &\text{for} \quad k = 1,2\ldots K \end{aligned}\right\} \tag{6.2.5.3}$$

At time t we use the following form of (6.2.5.2) in order to calculate the posterior $p_t^{(k)}$:

$$p_t^{(k)} \propto L_t^{(k)} \cdot p_{t-1}^{(k)*} \tag{6.2.5.4}$$

We then transform $p_t^{(k)}$ into $p_t^{(k)*}$ and the process continues in this way. Note at this point that the set of $p_t^{(k)*}$ sum to unity over all k:

$$\begin{aligned} \Sigma p_t^{(k)*} &= \Sigma p_L + \Sigma p_{t-1}^{(k)} - Kp_L\ \Sigma p_{t-1}^{(k)} \\ &= Kp_L + 1 - Kp_L = 1 \end{aligned}$$

In the following table we use $p_L = .0001$ and $K = 11$, i.e.

$$p_t^{(k)*} = .0001 + .9989\ p_t^{(k)}$$

in order to indicate the effect of this transformation on $p_t^{(k)}$ for different values of $p_t^{(k)}$:

TABLE 6.12

$p_t^{(k)}$	$p_t^{(k)*}$	$(p_t^{(k)*}/p_t^{(k)})$
.000001	.0001009989	100.9989
.00001	.000109989	10.9989
.0001	.00019989	1.9989
.001	.0010989	1.0989
.01	.010089	1.0089
.1	.09999	.9999
.2	.19988	.99945
.4	.39966	.99915
.8	.79922	.999025

It can be seen that high $p_t^{(k)}$ are virtually unaffected by the transformation but low $p_t^{(k)}$ are greatly increased. Table 6.11 shows the results produced by FRM using p_L = .0001 and is directly comparable with Table 6.10 where $p_L = 0$. Note that at time $t = 100$ $p_t^{(k)}$ and $\hat{V}_{\varepsilon,t}$ for both $p_L = 0$ and $p_L = .0001$ are almost identical to two decimal places illustrating the fact that p_L of this order makes no significant difference to FRM as long as V_ε is relatively stable (in this case $V_\varepsilon = .0025$ for $t = 1,2,...100$). If however V_ε has a step change as in the present example from $t = 100$ onwards, then $p_L > 0$ enables FRM to respond faster. This is clear by direct comparison of Tables 6.10 and 6.11.

The results from an experiment investigating the effect of p_L on $\hat{V}_{\varepsilon,t}$ for a large range of p_L are shown in Table 6.13. Based on this and similar experiments using different realisations of the C6 Data type it has been found that $p_L \leq .001$ is not critical to

t	$\rho_L = 0$	$\rho_L = .00001$	$\rho_L = .0001$	$\rho_L = .001$
2	.0071	.0071	.0071	.0072
4	.0034	.0034	.0034	.0035
6	.0020	.0020	.0020	.0021
8	.0023	.0023	.0024	.0025
10	.0020	.0020	.0020	.0021
12	.0015	.0015	.0015	.0015
14	.0019	.0019	.0019	.0021
16	.0023	.0023	.0023	.0026
18	.0027	.0027	.0027	.0032
20	.0026	.0026	.0026	.0029
22	.0024	.0024	.0024	.0025
24	.0022	.0022	.0022	.0023
26	.0019	.0019	.0019	.0020
28	.0017	.0017	.0017	.0017
30	.0018	.0018	.0018	.0020
32	.0017	.0017	.0017	.0017
34	.0015	.0015	.0015	.0015
36	.0015	.0015	.0016	.0016
38	.0018	.0018	.0019	.0033
40	.0018	.0019	.0025	.0063
42	.0019	.0020	.0022	.0039
44	.0018	.0018	.0019	.0024
46	.0016	.0016	.0016	.0017
48	.0015	.0015	.0015	.0016
50	.0018	.0018	.0018	.0022
52	.0018	.0018	.0018	.0021
54	.0020	.0020	.0020	.0024
56	.0020	.0020	.0021	.0024
58	.0021	.0021	.0021	.0024
60	.0021	.0021	.0021	.0024
62	.0019	.0019	.0020	.0023
64	.0020	.0020	.0020	.0020
66	.0019	.0019	.0019	.0022
68	.0020	.0020	.0021	.0019
70	.0023	.0023	.0024	.0023
72	.0022	.0022	.0022	.0031
74	.0021	.0021	.0021	.0025
76	.0020	.0020	.0020	.0022
78	.0020	.0020	.0020	.0020
80	.0020	.0020	.0020	.0021
82	.0021	.0021	.0022	.0020
84	.0020	.0020	.0021	.0025
86	.0021	.0021	.0021	.0021
88	.0021	.0021	.0021	.0022
90	.0021	.0021	.0021	.0021
92	.0020	.0020	.0020	.0022
94	.0021	.0021	.0022	.0021
96	.0021	.0021	.0021	.0025
98	.0020	.0020	.0021	.0023
100	.0020	.0020	.0020	.0021

Table 6.13

$\hat{V}_{\varepsilon,t}$ produced by FRM when applied to CG Data with $V_\varepsilon = \begin{cases} .0025 & t \le 100 \\ .0050 & t > 100 \end{cases}$

102	.0021	.0021	.0023	.0034
104	.0022	.0023	.0026	.0047
106	.0022	.0022	.0023	.0031
108	.0022	.0022	.0023	.0028
110	.0022	.0022	.0023	.0027
112	.0022	.0022	.0022	.0024
114	.0022	.0022	.0023	.0026
116	.0023	.0023	.0024	.0033
118	.0024	.0025	.0029	.0054
120	.0025	.0027	.0038	.0073
122	.0026	.0029	.0048	.0087
124	.0028	.0044	.0090	.0111
126	.0028	.0033	.0058	.0081
128	.0028	.0032	.0049	.0068
130	.0029	.0043	.0077	.0092
132	.0030	.0042	.0069	.0082
134	.0030	.0039	.0060	.0073
136	.0030	.0036	.0053	.0065
138	.0030	.0034	.0046	.0059
140	.0029	.0031	.0038	.0047
142	.0030	.0031	.0037	.0046
144	.0031	.0033	.0041	.0051
146	.0030	.0031	.0035	.0042
148	.0029	.0030	.0032	.0036
150	.0030	.0031	.0035	.0042
152	.0030	.0030	.0031	.0034
154	.0031	.0031	.0034	.0042
156	.0031	.0031	.0034	.0041
158	.0032	.0033	.0038	.0049
160	.0031	.0032	.0036	.0045
162	.0032	.0032	.0036	.0044
164	.0032	.0034	.0039	.0049
166	.0033	.0034	.0039	.0048
168	.0032	.0033	.0036	.0042
170	.0032	.0032	.0035	.0040
172	.0032	.0032	.0033	.0036
174	.0031	.0031	.0032	.0033
176	.0032	.0032	.0033	.0037
178	.0032	.0033	.0036	.0045
180	.0033	.0033	.0036	.0043
182	.0032	.0032	.0034	.0038
184	.0033	.0033	.0035	.0041
186	.0032	.0032	.0033	.0036
188	.0032	.0032	.0033	.0037
190	.0032	.0033	.0035	.0045
192	.0033	.0034	.0039	.0056
194	.0033	.0035	.0041	.0059
196	.0034	.0036	.0045	.0063
198	.0034	.0035	.0042	.0054
200	.0034	.0036	.0043	.0055

Table 6.13 cont.

$\hat{V}_{\varepsilon,t}$ and a choice of $p_L = .0001$ produces consistently a good balance in the trade off between $\hat{V}_{\varepsilon,t}$ instability and speed of response when V_ε changes abruptly. Note again that prior to $t = 100$ $\hat{V}_{\varepsilon,t}$ for all p_L in Table 6.11 are almost identical illustrating our previous statement that p_L does not affect $\hat{V}_{\varepsilon,t}$ unless a genuine change has taken place in the underlying level of the observation noise variance.

6.2.6. Illustration of FRM on 3MData and B/J Data

Consider first 3M Data described earlier in Section 6.2.1. It is known with hindsight that the final posterior estimate of V_ε is of the order of .0015 and therefore we have chosen an initial crude estimate $V_o = .0060$ for the construction of a V range with $K = 11$. The on line estimates $\hat{V}_{\varepsilon,t}$ are tabulated in Table 6.14, and the one step ahead forecasts are graphed in Figure 6.1. From Figure 6.1 it can be seen that 3M Data contain a period of increased variability starting from $t = 50$. This time period corresponds to November 1973 when the first world oil crisis of the 70's followed a relatively long period of stability and resulted in greater uncertainty giving rise to the increased variability in the data. The on line estimates given in Table 6.14 reflect this increased observation noise variance with a final $\hat{V}_{\varepsilon,t}$ estimate which is approximately 30% higher than the estimates produced during the quiet period before $t = 50$.

Finally in Table 6.15 and Figure 6.2 we show (i) the on line estimates $\hat{V}_{\varepsilon,t}$ and (ii) the one step ahead forecasts, respectively, produced by FRM when applied to B/J Data. The V range was

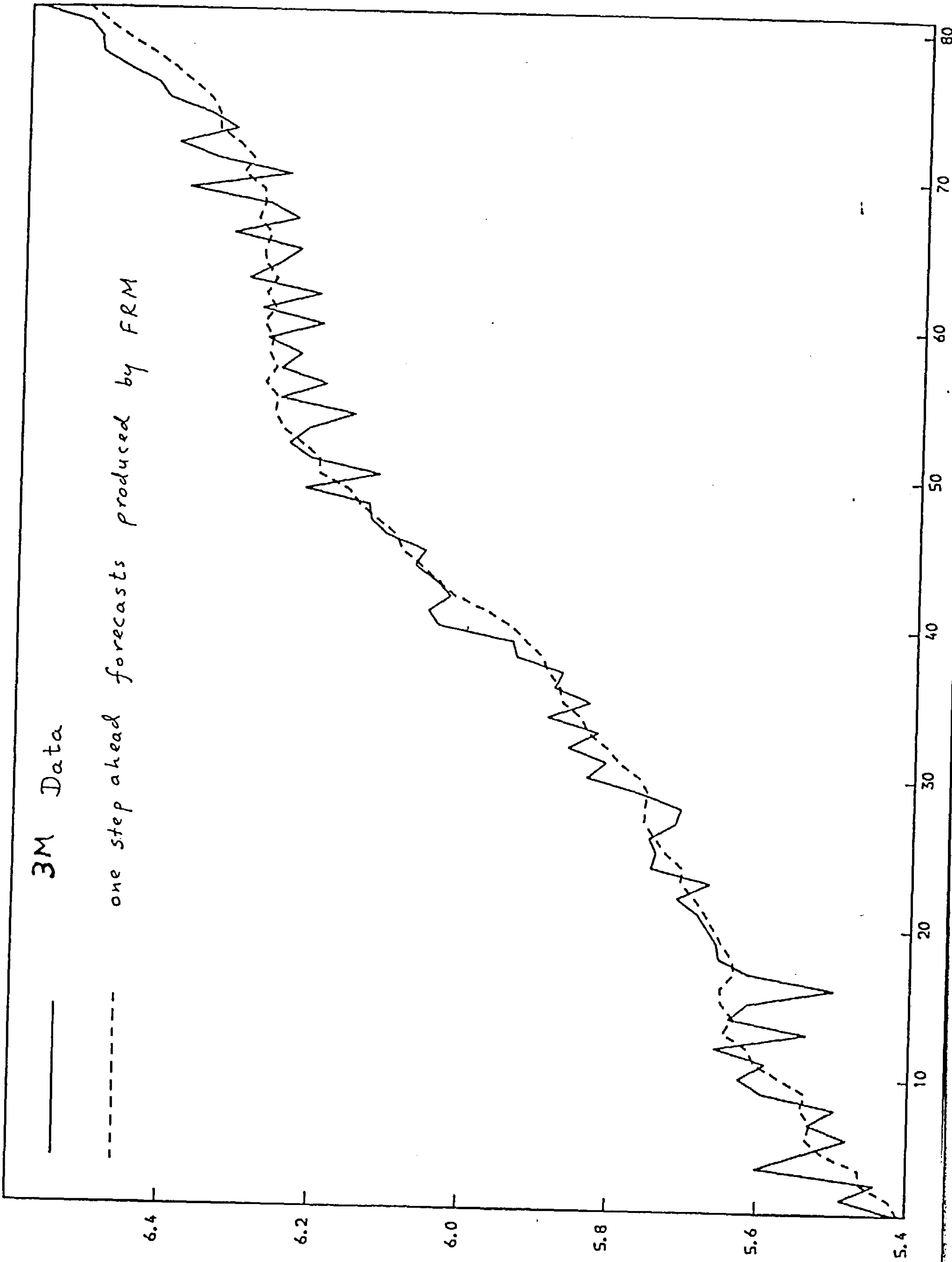
t	$\hat{V}_{\varepsilon,t}$
1	.0091
2	.0070
3	.0047
4	.0071
5	.0051
6	.0042
7	.0032
8	.0027
9	.0025
10	.0024
11	.0020
12	.0019
13	.0027
14	.0023
15	.0021
16	.0032
17	.0028
18	.0025
19	.0021
20	.0019
21	.0017
22	.0015
23	.0015
24	.0014
25	.0013
26	.0012
27	.0012
28	.0012
29	.0011
30	.0013
31	.0012
32	.0012
33	.0012
34	.0012
35	.0012
36	.0011
37	.0011
38	.0010
39	.0010
40	.0012

t	$\hat{V}_{\varepsilon,t}$
41	.0013
42	.0013
43	.0012
44	.0012
45	.0011
46	.0011
47	.0011
48	.0010
49	.0011
50	.0012
51	.0011
52	.0011
53	.0011
54	.0013
55	.0012
56	.0013
57	.0013
58	.0013
59	.0012
60	.0013
61	.0013
62	.0014
63	.0014
64	.0013
65	.0013
66	.0013
67	.0014
68	.0013
69	.0015
70	.0015
71	.0015
72	.0016
73	.0016
74	.0016
75	.0016
76	.0016
77	.0016
78	.0017
79	.0016
80	.0016
81	.0016

Table 6.14

FRM applied to 3M Data

FIGURE 6.1

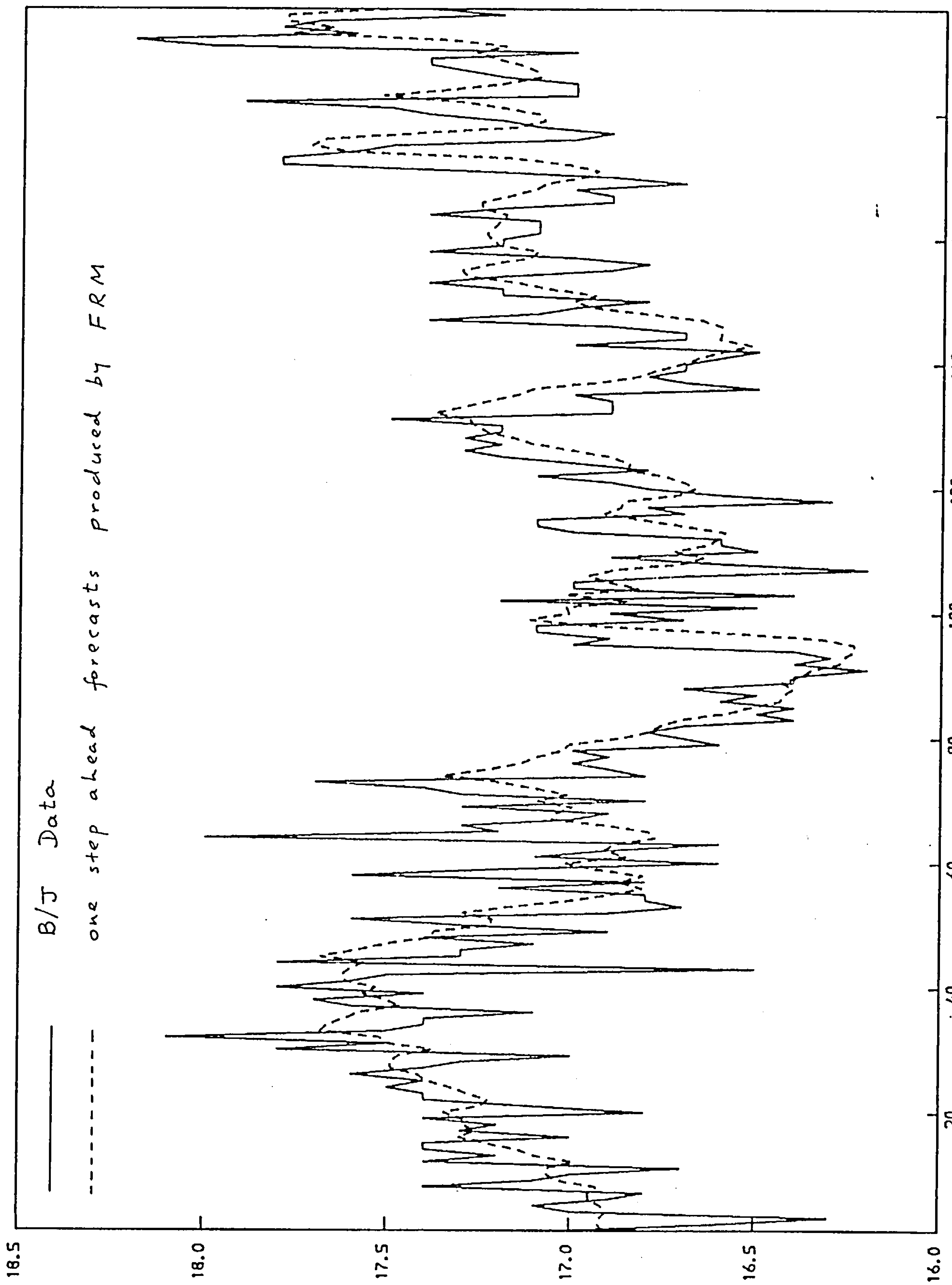

3M Data
one step ahead forecasts produced by FRM
6.4
6.2
6.0
5.8
5.6
5.4
10
20
30
40
50
60
70
80

t	$\hat{V}_{\varepsilon,t}$	t	$\hat{V}_{\varepsilon,t}$	t	$\hat{V}_{\varepsilon,t}$	t	$\hat{V}_{\varepsilon,t}$
1	.279	51	.056	101	.059	151	.058
2	.206	52	.058	102	.059	152	.057
3	.294	53	.061	103	.062	153	.057
4	.198	54	.060	104	.061	154	.056
5	.149	55	.058	105	.060	155	.057
6	.113	56	.060	106	.060	156	.058
7	.094	57	.058	107	.064	157	.058
8	.105	58	.065	108	.063	158	.058
9	.090	59	.064	109	.063	159	.057
10	.079	60	.065	110	.062	160	.056
11	.080	61	.065	111	.061	161	.055
12	.084	62	.063	112	.060	162	.055
13	.076	63	.063	113	.061	163	.054
14	.074	64	.062	114	.062	164	.054
15	.069	65	.064	115	.062	165	.053
16	.069	66	.064	116	.061	166	.054
17	.063	67	.063	117	.060	167	.054
18	.059	68	.061	118	.063	168	.053
19	.056	69	.061	119	.062	169	.054
20	.065	70	.061	120	.061	170	.053
21	.062	71	.061	121	.061	171	.054
22	.060	72	.061	122	.061	172	.056
23	.057	73	.064	123	.060	173	.057
24	.055	74	.068	124	.059	174	.056
25	.052	75	.068	125	.059	175	.056
26	.051	76	.067	126	.059	176	.058
27	.049	77	.066	127	.058	177	.059
28	.048	78	.064	128	.058	178	.058
29	.052	79	.066	129	.057	179	.057
30	.055	80	.065	130	.056	180	.057
31	.053	81	.063	131	.055	181	.057
32	.060	82	.062	132	.057	182	.059
33	.058	83	.062	133	.058	183	.061
34	.058	84	.061	134	.059	184	.061
35	.057	85	.059	135	.058	185	.060
36	.061	86	.058	136	.061	186	.060
37	.059	87	.057	137	.061	187	.059
38	.058	88	.057	138	.060	188	.059
39	.057	89	.056	139	.059	189	.058
40	.056	90	.055	140	.058	190	.058
41	.054	91	.054	141	.057	191	.062
42	.053	92	.053	142	.056	192	.064
43	.052	93	.052	143	.058	193	.064
44	.051	94	.051	144	.057	194	.063
45	.053	95	.055	145	.056	195	.062
46	.053	96	.057	146	.056	196	.064
47	.055	97	.058	147	.060	197	.063
48	.053	98	.057	148	.060		
49	.057	99	.058	149	.059		
50	.055	100	.057	150	.058		

Table 6.15

FRM applied to B/J Data

FIGURE 6.2


B/J Data
one step ahead forecasts produced by FRM
18.5
18.0
17.5
17.0
16.5
16.0
20
40
60
80
100
120
140
160
180

constructed using $V_0 = .160$, that is overestimating the final V_ε estimates (known with hindsight to be of the order of .065) by a factor of four. The final variance estimate produced by FRM is in fact $\hat{V}_{\varepsilon,197} = .063$ which is very close to that reported by Leonard and Harrison (L/H) [30]. L/H use a continuous on line estimation method designed only for the special case of a steady state process (not allowing any discontinuities) with constant variances. They report a final $\hat{V}_{\varepsilon,t}$ posterior of .066 and further details of their method will be given in Chapter 7.

6.3. The Variable Range Method (VRM)

This is an alternative estimation method to FRM with a significant computational advantage. The basic difference is that the prior information about the true noise variance at time t is represented in terms of a variable range consisting of only two values. An important implication of such a variable range is that changes in the observation noise variance are not restricted to within the extreme values of the fixed V range in the FRM.

Suppose that posterior to y_{t-1} and just before y_t is known, we have a best on line estimate of the true variance $V_{\varepsilon,t-1}$, denoted by $\hat{V}_{\varepsilon,t-1}$. This estimate can be used to update the process parameters of the MSM along the lines described in Section 3.3 with the only difference that $\hat{V}_{\varepsilon,t-1}$ is used instead of $V_{\varepsilon,N}$ in order to construct $V_\varepsilon^{(j)}$, $V_\mu^{(j)}$ and $V_\beta^{(j)}$ according to Table 3.1. Hence we can summarise all the MSM information prior to y_t in terms of (i) and (ii) below:

$$\text{(i)} \quad \left(\begin{matrix} \mu_{t-1} \\ \\ \beta_{t-1} \end{matrix} \;\middle|\; M_{t-1}^{(i)}, D_{t-1} \right) \sim N \left(\begin{matrix} m_{t-1}^{(i)} \\ \\ b_{t-1}^{(i)} \end{matrix} \; ; \; \underset{\sim}{C}_{t-1}^{(i)} \right) \qquad (6.3.1)$$

where the interpretation of the above is the same with that given for the FRM

$$\text{and (ii)} \quad p_{t-1}^{(i)} = \left. \begin{array}{l} \text{posterior probability of being in} \\ \text{state } i \text{ at time } t-1 \\ \text{(as in FRM and MSM)} \end{array} \right\} \qquad (6.3.2)$$

The method now outlined is to produce an estimate $\hat{V}_{\varepsilon,t}$ given the actual observation y_t. Briefly, this is achieved by postulating two distributions for the random observation noise, with zero means and variances $V_{\varepsilon,t}^{(1)}$, $V_{\varepsilon,t}^{(2)}$ respectively and measuring the relative likelihood that y_t could have come from systems with these two variances. The new estimate $\hat{V}_{\varepsilon,t}$ will then be the weighted average of these two variances.

The selection of $V_{\varepsilon,t}^{(1)}$ and $V_{\varepsilon,t}^{(2)}$ has been constrained by choosing two constants c_1 and c_2 such that,

$$\left.\begin{aligned} V^{(1)}_{\varepsilon,t} &= c_1 \hat{V}_{\varepsilon,t-1} \\ V^{(2)}_{\varepsilon,t} &= c_2 \hat{V}_{\varepsilon,t-1} \end{aligned}\right\} \qquad (6.3.3)$$

where
$$\left.\begin{aligned} 0 < c_1 &< 1 \\ c_2 &> 1 \end{aligned}\right\} \qquad (6.3.4)$$

The choice of c_1 and c_2 is discussed in Section 6.4 but as an example a possible choice is $c_1 = 0.5$ and $c_2 = 2$.

Another constraint has been introduced to reflect the need that the weighted average of the two variances $V^{(1)}_{\varepsilon,t}$ and $V^{(2)}_{\varepsilon,t}$ prior to observing y_t should in fact equal our best estimate of the variance posterior to y_{t-1}, i.e. $\hat{V}_{\varepsilon,t-1}$.

If we therefore let $p^{(1)}$, $p^{(2)}$ be the prior probabilities associated with $V^{(1)}_{\varepsilon,t}$, $V^{(2)}_{\varepsilon,t}$ i.e.

$$\left.\begin{aligned} p^{(1)} &= p(V^{(1)}_{\varepsilon,t} \mid D_{t-1}) \\ p^{(2)} &= p(V^{(2)}_{\varepsilon,t} \mid D_{t-1}) \end{aligned}\right\} \qquad (6.3.5)$$

then the following must hold:

$$\left.\begin{aligned} p^{(1)} + p^{(2)} &= 1 \\ p^{(1)}V^{(1)}_{\varepsilon,t} + p^{(2)}V^{(2)}_{\varepsilon,t} &= \hat{V}_{\varepsilon,t-1} \end{aligned}\right\}$$

Substituting for $V^{(1)}_{\varepsilon,t}$, $V^{(2)}_{\varepsilon,t}$ from (6.3.3) and solving for $p^{(1)}$ and $p^{(2)}$ we get:

$$\left.\begin{aligned} p^{(1)} &= (c_2 - 1) \,/\, (c_2 - c_1) \\ p^{(2)} &= (1 - c_1) \,/\, (c_2 - c_1) \end{aligned}\right\} \qquad (6.3.6)$$

As soon as y_t becomes available it is viewed by VRM as having come from one of two sets of predictive distributions each of which corresponds to a system state transition $i \to j$ and either $V^{(1)}_{\varepsilon,t}$ or $V^{(2)}_{\varepsilon,t}$:

$$(y_t \mid V^{(k)}_{\varepsilon,t} , M^{(ij)} , D_{t-1}) \sim N(\hat{y}^{(ij)} , \hat{Y}^{(ijk)}) \qquad (6.3.7)$$

$\hat{Y}^{(ijk)}$ for $k = 1,2$ is associated with $V^{(1)}_{\varepsilon,t}$ and $V^{(2)}_{\varepsilon,t}$ respectively. The likelihoods of y_t having been generated by one of these predictive distributions are then given by:

$$L(y_t \mid V^{(k)}_{\varepsilon,t} , M^{(ij)} , D_{t-1}) \propto \left[\hat{Y}^{(ijk)}\right]^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}\left[e^{(ij)}\right]^2 / \hat{Y}^{(ijk)}\right\} \qquad (6.3.8)$$

where $e^{(ij)}$ and $\hat{Y}^{(ijk)}$ for $k = 1,2$ and $i,j = 1,2,3,4$ are the same as for FRM.

If we now let $p_t^{(1)}$, $p_t^{(2)}$ be the posterior probabilities for $V_{\varepsilon,t}^{(1)}$, $V_{\varepsilon,t}^{(2)}$ respectively then we have:

$$p_t^{(k)} = p(V_{\varepsilon,t}^{(k)} \mid y_t , D_{t-1}) \propto p(y_t \mid V_{\varepsilon,t}^{(k)} , D_{t-1}) \cdot p(V_{\varepsilon,t}^{(k)} \mid D_{t-1})$$

or using (6.3.5)

$$p_t^{(k)} \propto p(y_t \mid V_{\varepsilon,t}^{(k)} , D_{t-1}) \cdot p^{(k)} \qquad (6.3.9)$$

and since $p^{(k)}$ is a known function of c_1 and c_2 (see (6.3.6.)) we only have to determine the first term of (6.3.9):

$$p(y_t \mid V_{\varepsilon,t}^{(k)} , D_{t-1}) = \sum_i \sum_j p(y_t \mid V_{\varepsilon,t}^{(k)} , M^{(ij)} , D_{t-1}) p(M^{(ij)} \mid V_{\varepsilon,t}^{(k)} , D_{t-1})$$

$$\propto \sum_i \sum_j L(y_t \mid V_{\varepsilon,t}^{(k)} , M^{(ij)} , D_{t-1}) \cdot p_{t-1}^{(i)} \cdot \pi^{(j)}$$

and therefore $p_t^{(k)}$ can be fully determined by substituting the known quantities of (6.3.8) and (6.3.2) for the first and second terms in the last expression.

The posterior estimate of the true observation noise variance at time t can now be calculated as the weighted average of the two postulated variances:

$$\hat{V}_{\varepsilon,t} = \sum_k p_t^{(k)} V_{\varepsilon,t}^{(k)} \tag{6.3.10}$$

At this point we have completed a whole cycle of the VRM procedure since $\hat{V}_{\varepsilon,t}$ can now be used to update the process parameters of the MSM and thus arrive at prior information for time $t+1$ in exactly the same form as given by (6.3.1) and (6.3.2). As soon as y_{t+1} becomes available the whole procedure can be repeated.

6.4. Illustration of the VRM

6.4.1. Foreword

In this section we will illustrate VRM numerically in a similar way as for FRM previously, and comparisons of the two procedures will be made. The different areas examined here are:

(i) specification of the variable V range in terms of the constants c_1 and c_2.

(ii) speed of initial convergence given an initial crude estimate of the noise variance at $t = 0$, $V_{\varepsilon,0}$

(iii) the effect of discontinuities such as outliers, growth changes and step changes on the on line estimates $\hat{V}_{\varepsilon,t}$.

(iv) the response of the method to a discontinuity in the observation noise variance V_ε.

Finally the method will be applied to 3M Data and B/J Data.

6.4.2. Specification of the variable V range

The variable range method described in 6.3 assumed that the prior information about the true noise variance at time t, can be represented in terms of a variable range $(V^{(1)}_{\varepsilon,t}\ ,\ V^{(2)}_{\varepsilon,t})$ consisting of only two values:

$$\left.\begin{aligned} V^{(1)}_{\varepsilon,t} &= c_1\ \hat{V}_{\varepsilon,t-1} \\ V^{(2)}_{\varepsilon,t} &= c_2\ \hat{V}_{\varepsilon,t-1} \end{aligned}\right\} \qquad (6.4.2.1)$$

where $\hat{V}_{\varepsilon,t-1}$ is our best on line estimate of the noise variance posterior to time t-1.

A more general formulation of the VRM,

$$\left.\begin{aligned} & V^{(k)}_{\varepsilon,t} = c_k\ \hat{V}_{\varepsilon,t-1} \\ & k = 1,2,3,\ldots K \\ & K > 2 \end{aligned}\right\} \qquad (6.4.2.2)$$

has also been fully tested but has lead to no significant improvement and the results are therefore not reported. Instead we will concentrate on the $K = 2$ case which produces a good response and is computationally very efficient. The CPU time required by the VRM with $K = 2$ is in fact less than one third that required by FRM with our recommended choice of $K = 11$.

Given $K = 2$ it follows that at time $t = 0$ prior to any observations, VRM requires values for c_1, c_2 and $\hat{V}_{\varepsilon,0}$. There are many ways one can choose the c's and we propose the following which has been found to work successfully:

$$\left.\begin{aligned} c_1 &= 1/\beta \\ c_2 &= \beta \\ \beta &> 1 \end{aligned}\right\} \qquad (6.4.2.3)$$

This choice ensures that the lower value in our range (i.e. $V^{(1)}_{\varepsilon,t}$) will always be positive and in addition VRM becomes a single parameter method. That is given β, c_1 and c_2 are determined from (6.4.2.3) and $p^{(1)}$, $p^{(2)}$ from (6.3.6):

$$\left.\begin{aligned} p^{(1)} &= \beta/(\beta+1) \\ p^{(2)} &= 1/(\beta+1) \end{aligned}\right\} \qquad (6.4.2.4)$$

The interpretation of β is simple. Since our best on line estimate posterior to time t (i.e. $\hat{V}_{\varepsilon,t}$) is going to be a linear combination of $V^{(1)}_{\varepsilon,t}$ and $V^{(2)}_{\varepsilon,t}$ (see equation (6.3.10)), then β is the maximum factor by which $\hat{V}_{\varepsilon,t}$ can differ from $\hat{V}_{\varepsilon,t-1}$.

Consider now the range $(V^{(1)}_{\varepsilon,t}\ ,\ V^{(2)}_{\varepsilon,t})$. It follows from (6.4.2.1) and (6.4.2.3) that:

$$V^{(2)}_{\varepsilon,t} \;=\; \beta^2 \;\; V^{(1)}_{\varepsilon,t}$$

and therefore as β increases so does the range. Consequently it is expected that larger β are likely to produce more responsive on line estimates $\hat{V}_{\varepsilon,t}$. This is confirmed by the results of Table 6.16, showing $\hat{V}_{\varepsilon,t}$ produced by VRM when applied to C6 Data and starting with an initial estimate of $\hat{V}_{\varepsilon,0} = .0050$. It can be seen that $\beta = 1.6$ converges to the true $V_\varepsilon = .0025$ level much faster than $\beta = 1.2$ but is also more unstable. At time $t = 118$ for example $\hat{V}_{\varepsilon,t}$ corresponding to $\beta = 1.6$ is .0019 while at $t = 130$ $\hat{V}_{\varepsilon,t} =$ .0038, that is the on line estimate has doubled in 12 time periods. From a large number of experiments including those to be described in the following three sections it has been established that $\beta = 1.4$ produces a consistently good response and this value of β can therefore be recommended as a trade off between stability and speed of response.

6.4.3. Speed of initial convergence

If $\hat{V}_{\varepsilon,0}$ is in large error from V_ε, say by a factor of 5, then our recommended value of $\beta = 1.4$ would take rather longer to converge than a value of $\beta = 1.6$, say, which was seen to be quite responsive. The reason why VRM with $\beta = 1.4$ is rather slow initially is because the two extremes of the range are apart by a factor of $\beta^2 = 1.96$ (i.e. $V^{(2)}_{\varepsilon,t} = 1.96\ V^{(1)}_{\varepsilon,t}$) in contrast with FRM where the extreme values of its range are apart by a factor of 7.59^2 (i.e.

Table 6.16

t	β=1.2	β=1.3	β=1.4	β=1.5	β=1.6
2	.0050	.0049	.0049	.0048	.0048
4	.0049	.0047	.0045	.0043	.0041
6	.0048	.0045	.0042	.0039	.0037
8	.0048	.0045	.0043	.0040	.0038
10	.0047	.0044	.0040	.0037	.0034
12	.0046	.0042	.0037	.0033	.0029
14	.0046	.0042	.0038	.0034	.0031
16	.0045	.0042	.0038	.0035	.0032
18	.0046	.0042	.0039	.0036	.0035
20	.0045	.0041	.0037	.0035	.0033
22	.0044	.0039	.0035	.0032	.0029
24	.0043	.0038	.0034	.0030	.0027
26	.0042	.0036	.0031	.0027	.0024
28	.0041	.0035	.0029	.0024	.0021
30	.0041	.0034	.0028	.0024	.0021
32	.0040	.0033	.0027	.0022	.0019
34	.0039	.0031	.0025	.0020	.0017
36	.0039	.0031	.0024	.0020	.0017
38	.0040	.0033	.0028	.0023	.0020
40	.0041	.0036	.0031	.0026	.0022
42	.0041	.0035	.0030	.0026	.0022
44	.0040	.0034	.0029	.0025	.0021
46	.0039	.0033	.0027	.0023	.0018
48	.0039	.0032	.0026	.0021	.0017
50	.0039	.0032	.0027	.0023	.0020
52	.0038	.0032	.0027	.0023	.0020
54	.0038	.0032	.0027	.0024	.0022
56	.0038	.0031	.0027	.0024	.0023
58	.0038	.0031	.0027	.0025	.0023
60	.0037	.0030	.0026	.0024	.0023
62	.0036	.0029	.0024	.0022	.0020
64	.0036	.0029	.0025	.0023	.0021
66	.0035	.0028	.0023	.0020	.0019
68	.0035	.0028	.0025	.0023	.0022
70	.0036	.0030	.0028	.0028	.0029
72	.0035	.0029	.0026	.0025	.0025
74	.0035	.0028	.0024	.0023	.0022
76	.0034	.0027	.0023	.0020	.0019
78	.0034	.0026	.0023	.0021	.0020
80	.0033	.0026	.0022	.0019	.0018
82	.0033	.0027	.0024	.0023	.0023
84	.0033	.0026	.0022	.0021	.0020
86	.0032	.0026	.0023	.0021	.0021
88	.0032	.0025	.0022	.0021	.0020
90	.0032	.0026	.0023	.0022	.0021
92	.0031	.0025	.0022	.0020	.0019
94	.0032	.0026	.0024	.0023	.0023
96	.0031	.0025	.0023	.0022	.0022
98	.0031	.0024	.0021	.0020	.0019
100	.0030	.0023	.0020	.0018	.0017

$\hat{V}_{E,t}$ produced by VRM when applied to C6 Data with $\hat{V}_{E,0} = .0050$

Table 6.16 cont.

t	β=1.2	β=1.3	β=1.4	β=1.5	β=1.6
102	.0030	.0024	.0022	.0021	.0021
104	.0031	.0025	.0023	.0023	.0024
106	.0030	.0024	.0022	.0021	.0021
108	.0030	.0024	.0022	.0021	.0020
110	.0029	.0023	.0021	.0020	.0019
112	.0029	.0022	.0019	.0018	.0017
114	.0028	.0022	.0019	.0018	.0017
116	.0029	.0023	.0020	.0020	.0019
118	.0029	.0024	.0023	.0023	.0024
120	.0030	.0026	.0026	.0027	.0029
122	.0031	.0027	.0027	.0030	.0033
124	.0032	.0029	.0031	.0035	.0039
126	.0031	.0028	.0029	.0031	.0034
128	.0031	.0027	.0028	.0030	.0031
130	.0032	.0029	.0031	.0034	.0038
132	.0031	.0029	.0031	.0034	.0036
134	.0031	.0029	.0030	.0032	.0033
136	.0031	.0028	.0029	.0030	.0031
138	.0030	.0027	.0027	.0028	.0028
140	.0030	.0026	.0025	.0025	.0024
142	.0029	.0026	.0025	.0024	.0024
144	.0029	.0026	.0026	.0025	.0025
146	.0029	.0025	.0024	.0023	.0022
148	.0028	.0024	.0022	.0021	.0019
150	.0028	.0024	.0023	.0022	.0021
152	.0028	.0023	.0021	.0020	.0018
154	.0028	.0024	.0022	.0021	.0020
156	.0027	.0023	.0022	.0021	.0020
158	.0028	.0024	.0024	.0023	.0023
160	.0027	.0024	.0022	.0022	.0022
162	.0027	.0023	.0022	.0021	.0021
164	.0028	.0024	.0024	.0024	.0024
166	.0027	.0024	.0024	.0024	.0024
168	.0027	.0023	.0022	.0021	.0020
170	.0026	.0022	.0021	.0020	.0019
172	.0026	.0021	.0020	.0018	.0017
174	.0025	.0021	.0018	.0016	.0015
176	.0025	.0021	.0019	.0017	.0016
178	.0026	.0022	.0021	.0021	.0020
180	.0026	.0022	.0021	.0021	.0021
182	.0025	.0021	.0019	.0018	.0018
184	.0025	.0021	.0020	.0019	.0019
186	.0025	.0020	.0018	.0017	.0016
188	.0024	.0020	.0018	.0017	.0016
190	.0025	.0021	.0020	.0019	.0018
192	.0026	.0022	.0022	.0021	.0021
194	.0026	.0023	.0023	.0023	.0024
196	.0027	.0024	.0025	.0026	.0027
198	.0026	.0024	.0024	.0025	.0025
200	.0027	.0024	.0025	.0025	.0027

$V_\varepsilon^{(11)} = 58\ V_\varepsilon^{(1)}$ approximately).

It is therefore proposed to operate with a FRM during the initial stages until an arbitrary point in time, say $t = 20$. At this point $\hat{V}_{\varepsilon,20}$ is a much better estimate of V_ε than our initial crude estimate and hence the VRM can now be used (in view of its computational advantage) with $\beta = 1.4$ and an initial estimate $\hat{V}_{\varepsilon,0}$ equal to $\hat{V}_{\varepsilon,20}$. This has been tested on over fifty realisations of the C6 Data type and $\hat{V}_{\varepsilon,20}$ in all cases was never in error from V_ε by more than a factor of 2. This was true even when our initial crude estimate was in error by unrealistically large factors such as ten or larger. The fast FRM initial convergence has in fact already been illustrated in Section 6.2.2.(see Tables 6.4 and 6.5) which concluded that the effect of our initial crude estimate dies away very quickly and $\hat{V}_{\varepsilon,t}$ is then wholly determined by the actual data.

6.4.4. Effect of process discontinuities

The experiment described earlier in 6.2.4 has been repeated the only difference being that VRM ($\beta = 1.4$, $\hat{V}_{\varepsilon,0} = .0100$) has been used instead of FRM ($\kappa = 11$, $V_0 = .0100$). Table 6.17 is directly comparable with Table 6.8 showing $\hat{V}_{\varepsilon,t}$ for the different types and sizes of discontinuities introduced on C6 Data at time $t = 61$.

Comparing $\hat{V}_{\varepsilon,t,*}$ with all other $\hat{V}_{\varepsilon,t}$ leads to the same conclusion as in FRM, that is, the effect of discontinuities is only a small overestimation of $\hat{V}_{\varepsilon,t,*}$, the largest difference $(\hat{V}_{\varepsilon,t} - \hat{V}_{\varepsilon,t,*})$

Table 6.17

$\hat{V}_{\varepsilon,t} \times 10^4$

t	$V_{\varepsilon,t,*} \times 10^4$	OUTLIER of size 4σ	10σ	20σ	GROWTH CHANGE of size $\frac{1}{2}\sigma$	1σ	1.5σ	STEP CHANGE of size 2σ	5σ	10σ
51	26	26	26	26	26	26	26	26	26	26
52	27	27	27	27	27	27	27	27	27	27
53	28	28	28	28	28	28	28	28	28	28
54	27	27	27	27	27	27	27	27	27	27
55	26	26	26	26	26	26	26	26	26	26
56	27	27	27	27	27	27	27	27	27	27
57	26	26	26	26	26	26	26	26	26	26
58	27	27	27	27	27	27	27	27	27	27
59	27	27	27	27	27	27	27	27	27	27
60	26	26	26	26	26	26	26	26	26	26
61	25	27	26	31	26	27	28	27	26	26
62	24	26	25	30	25	27	31	29	26	26
63	24	25	25	29	24	27	34	31	25	25
64	25	26	26	30	28	29	34	30	26	26
65	24	25	25	28	29	30	34	29	26	25
66	23	24	24	27	30	30	33	28	25	24
67	25	26	26	29	34	30	33	29	25	26
68	25	26	25	29	33	30	33	29	26	26
69	25	26	26	29	35	30	33	29	26	26
70	28	29	29	32	35	32	35	31	28	29
71	27	28	27	30	34	31	34	30	27	27
72	26	27	27	30	34	30	33	30	27	27
73	25	26	26	29	33	30	32	30	26	26
74	24	25	25	28	32	29	31	28	25	25
75	24	24	24	27	31	28	30	27	24	24
76	23	23	23	26	30	27	29	26	23	23
77	23	23	23	25	29	27	29	27	23	23
78	23	23	23	25	29	27	29	26	23	23
79	22	22	22	24	28	26	28	25	22	23
80	22	22	22	24	28	25	27	25	22	22
81	25	25	25	27	30	28	30	28	25	25
82	24	25	25	26	29	27	29	28	25	25
83	23	24	24	26	28	26	28	27	24	24
84	22	23	23	25	27	25	27	26	23	23
85	23	23	23	25	28	26	28	26	23	23
86	23	23	23	25	27	25	27	26	23	23
87	23	23	23	25	27	25	27	26	23	23
88	22	23	23	24	26	25	26	25	23	23
89	24	24	24	25	28	26	27	26	24	24
90	23	23	23	24	27	25	26	25	23	23
91	22	22	22	23	26	24	25	24	22	22
92	22	22	22	23	25	24	25	24	22	22
93	22	22	22	23	25	24	25	24	22	22
94	24	24	24	25	27	26	27	26	24	24
95	23	23	23	25	26	25	26	25	23	24
96	23	23	23	24	26	25	26	25	23	23
97	22	22	22	23	25	24	25	24	22	22
98	21	21	21	22	24	23	24	23	21	21
99	21	21	21	22	23	22	23	22	21	21
100	20	20	20	21	23	21	22	22	20	20

$\hat{V}_{\varepsilon,t} \times 10^4$ produced by VRM when applied to C6 Data with discontinuities at $t = 61$

$\hat{V}_{\varepsilon,t,*} \times 10^4$ has been produced by VRM when applied to C6 Data without any discontinuity at $t = 61$

being of the order of 30%. This however is not critical to the performance of the system at points of discontinuities as illustrated with the following example using Figure 6.3.

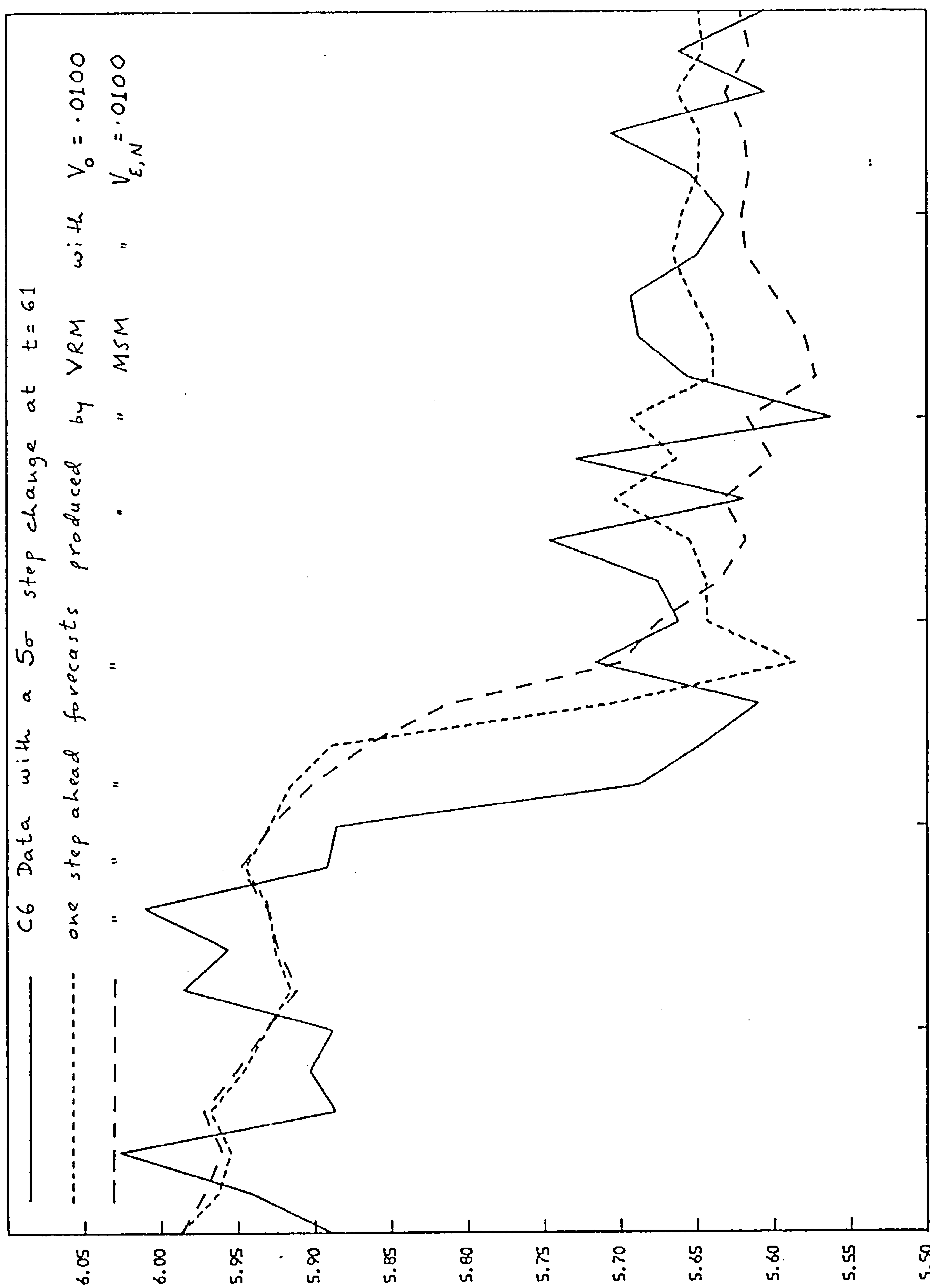
In the last section of Chapter 5 it was shown (using a purely deterministic set up) that for step changes of the order of 5σ if the MSM is used with $V_{\varepsilon,N} = 4V_{\varepsilon}$ (i.e. overestimating the true variance by a factor of four) the system level responds very slowly, undershoots and remains significantly biased for over 15 time periods. In Figure 6.3 we show the one step ahead forecast response of:

(i) The MSM operating with $V_{\varepsilon,N} = 4V_{\varepsilon} = .0100$ i.e. overestimating the true noise variance by a factor of 4,

(ii) The VRM producing on line estimates of the variance and having started with $\hat{V}_{\varepsilon,0} = V_{\varepsilon,N} = .0100$,

on C6 Data with a 5σ discontinuity in level (step change) at $t = 61$. It can be seen that before $t = 61$ the two sets of forecasts produced are very similar but after the step change at $t = 61$ VRM adapts very quickly to the new situation in contrast with MSM which responds slower and produces biased forecasts for a considerable period of time. This is a good example of the advantages of on line variance estimation and the risk one might be taking when operating the MSM with a fixed initially nominated estimate $V_{\varepsilon,N}$.

FIGURE 6.3


C6 Data with a 5σ step change at t = 61
one step ahead forecasts produced by VRM with V_o = ·0100
" " " " " " MSM " $V_{\varepsilon,N}$ = ·0100
6.05
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
5.60
5.55
5.50
50
55
60
65
70
75

6.4.5. Effect of a discontinuity in V_ε

As in 6.2.5 the effect of an abrupt change in the observation noise variance is investigated using C6 Data generated with:

$$V_\varepsilon = \begin{cases} .0025 & t \leq 100 \\ .0050 & t > 100 \end{cases} \qquad (6.4.5.1)$$

The on line estimates $\hat{V}_{\varepsilon,t}$ produced by VRM using $\beta = 1.2, 1.3, 1.4, 1.5, 1.6$ are tabulated in Table 6.18 for time $t = 102, 104, 106, \ldots 200$. Table 6.18 for $t = 1,2,\ldots 100$, is not given since it is identical to Table 6.16 earlier which has been produced using C6 Data generated with $V_\varepsilon = .0025$ for $t = 1,2,\ldots 200$.

The V_ε discontinuity can be considered to be "identified" by VRM at a point in time t say, if $\hat{V}_{\varepsilon,t}$ from Table 6.18 is approximately twice as large as $\hat{V}_{\varepsilon,t}$ from Table 6.16. For our recommended value of $\beta = 1.4$ this happens around time $t = 160$ since $\hat{V}_{\varepsilon,t}$ (Table 6.18) = .0040 compared with $\hat{V}_{\varepsilon,t}$ (Table 6.16) = .0022. Note that although values of $\beta > 1.4$ appear to respond faster to the V_ε change a careful examination of Tables 6.18 and 6.16 shows that this is only the result of the fact that values of $\beta > 1.4$ are generally more unstable even when V_ε does not have a discontinuity (as is the case in Table 6.16).

Table 6.18

t	β=1.2	β=1.3	β=1.4	β=1.5	β=1.6
102	.0031	.0025	.0023	.0022	.0021
104	.0032	.0027	.0026	.0025	.0026
106	.0032	.0027	.0025	.0024	.0024
108	.0032	.0026	.0024	.0024	.0024
110	.0032	.0027	.0025	.0025	.0025
112	.0031	.0026	.0024	.0023	.0022
114	.0031	.0026	.0025	.0025	.0025
116	.0032	.0028	.0027	.0028	.0029
118	.0033	.0030	.0031	.0033	.0037
120	.0035	.0032	.0034	.0039	.0045
122	.0035	.0034	.0037	.0044	.0052
124	.0037	.0037	.0044	.0054	.0066
126	.0037	.0036	.0041	.0048	.0057
128	.0037	.0036	.0041	.0047	.0054
130	.0038	.0039	.0046	.0056	.0066
132	.0038	.0040	.0047	.0056	.0064
134	.0038	.0040	.0046	.0053	.0060
136	.0038	.0039	.0045	.0051	.0056
138	.0038	.0039	.0043	.0048	.0052
140	.0037	.0037	.0040	.0043	.0045
142	.0037	.0037	.0040	.0043	.0044
144	.0038	.0038	.0042	.0045	.0047
146	.0037	.0037	.0039	.0041	.0041
148	.0036	.0036	.0037	.0037	.0036
150	.0037	.0037	.0039	.0040	.0039
152	.0036	.0035	.0036	.0036	.0034
154	.0037	.0037	.0038	.0039	.0039
156	.0037	.0036	.0038	.0039	.0038
158	.0038	.0039	.0042	.0044	.0046
160	.0037	.0038	.0040	.0041	.0042
162	.0037	.0038	.0040	.0041	.0041
164	.0039	.0040	.0043	.0045	.0047
166	.0039	.0040	.0043	.0046	.0047
168	.0038	.0038	.0040	.0040	.0040
170	.0037	.0037	.0038	.0038	.0037
172	.0037	.0036	.0036	.0035	.0033
174	.0036	.0035	.0034	.0032	.0029
176	.0036	.0035	.0035	.0034	.0032
178	.0038	.0038	.0040	.0040	.0041
180	.0038	.0038	.0040	.0041	.0041
182	.0037	.0037	.0037	.0036	.0035
184	.0037	.0037	.0038	.0038	.0038
186	.0036	.0036	.0035	.0034	.0032
188	.0036	.0036	.0035	.0034	.0033
190	.0037	.0037	.0037	.0037	.0035
192	.0038	.0039	.0041	.0042	.0043
194	.0039	.0041	.0044	.0046	.0048
196	.0040	.0043	.0048	.0051	.0054
198	.0040	.0043	.0046	.0049	.0051
200	.0041	.0044	.0047	.0050	.0053

$\hat{V}_{\varepsilon,t}$ produced by VRM when applied to C6 Data

with $V_\varepsilon = \begin{cases} .0025 & t \leq 100 \\ .0050 & t > 100 \end{cases}$

6.4.6. Illustration of VRM on 3M Data and B/J Data

Tables 6.19 and 6.20 show the on line $\hat{V}_{\varepsilon,t}$ estimates produced by VRM with $\beta = 1.4$ when applied to 3M Data and B/J Data respectively. The procedure suggested in 6.4.3 has been used to select $\hat{V}_{\varepsilon,0}$ for VRM. That is $\hat{V}_{\varepsilon,0}$ for 3M Data and B/J Data was taken equal to .0019 and .065 respectively, which are the on line $t = 20$ variance estimates produced by FRM as shown in Tables 6.14 and 6.15.

Comparing VRM and FRM on these two sets of data it can be seen that the on line $\hat{V}_{\varepsilon,t}$ estimates produced by the two methods are not significantly different, the largest difference $\hat{V}_{\varepsilon,t}$(FRM) − $\hat{V}_{\varepsilon,t}$(VRM) at any point in time (after $t = 20$, thus allowing some initial convergence) is of the order of 30%. This sort of difference is not critical to the forecasting performance of the system and the one step ahead forecast response graphs produced by VRM are not in fact given since they are indistinguishable from figures 6.1 and 6.2 given earlier.

6.5. Concluding remarks

Two methods of on line observation noise variance estimation have been proposed.

FRM requires a fixed range to be specified initially in terms of V_o, K and c. The effect of V_o was shown to be insignificant and the values of $K = 11$ and $c = 1.5$ were recommended on grounds of

t	$\hat{V}_{\varepsilon,t}$	t	$\hat{V}_{\varepsilon,t}$
1	.0017	41	.0016
2	.0017	42	.0015
3	.0016	43	.0015
4	.0019	44	.0014
5	.0018	45	.0014
6	.0018	46	.0013
7	.0018	47	.0013
8	.0018	48	.0012
9	.0018	49	.0013
10	.0018	50	.0014
11	.0017	51	.0014
12	.0017	52	.0013
13	.0019	53	.0013
14	.0019	54	.0015
15	.0018	55	.0014
16	.0020	56	.0016
17	.0020	57	.0015
18	.0019	58	.0015
19	.0019	59	.0014
20	.0018	60	.0016
21	.0017	61	.0015
22	.0017	62	.0016
23	.0016	63	.0016
24	.0016	64	.0015
25	.0016	65	.0015
26	.0015	66	.0015
27	.0015	67	.0016
28	.0015	68	.0015
29	.0014	69	.0017
30	.0015	70	.0018
31	.0015	71	.0018
32	.0015	72	.0019
33	.0014	73	.0018
34	.0014	74	.0017
35	.0014	75	.0018
36	.0014	76	.0018
37	.0013	77	.0018
38	.0013	78	.0019
39	.0013	79	.0019
40	.0015	80	.0018
		81	.0018

Table 6.19

$\hat{V}_{\varepsilon,t}$ produced by VRM ($\beta = 1.4$; $\hat{V}_{\varepsilon,0} = .0019$)

when applied to 3M Data

Table 6.20

t	$\hat{V}_{\varepsilon,t}$	t	$\hat{V}_{\varepsilon,t}$	t	$\hat{V}_{\varepsilon,t}$	t	$\hat{V}_{\varepsilon,t}$
1	.062	51	.058	101	.055	151	.056
2	.064	52	.060	102	.056	152	.055
3	.074	53	.064	103	.061	153	.054
4	.072	54	.063	104	.060	154	.052
5	.070	55	.061	105	.058	155	.055
6	.067	56	.062	106	.057	156	.058
7	.065	57	.060	107	.065	157	.056
8	.069	58	.069	108	.063	158	.056
9	.066	59	.068	109	.062	159	.054
10	.064	60	.070	110	.061	160	.052
11	.066	61	.069	111	.059	161	.050
12	.068	62	.066	112	.056	162	.049
13	.065	63	.066	113	.059	163	.047
14	.064	64	.064	114	.061	164	.046
15	.062	65	.067	115	.061	165	.045
16	.063	66	.067	116	.060	166	.047
17	.060	67	.065	117	.058	167	.047
18	.058	68	.063	118	.064	168	.046
19	.056	69	.063	119	.062	169	.048
20	.063	70	.063	120	.060	170	.046
21	.061	71	.063	121	.058	171	.048
22	.059	72	.063	122	.060	172	.054
23	.057	73	.068	123	.057	173	.055
24	.056	74	.074	124	.056	174	.054
25	.053	75	.074	125	.056	175	.053
26	.052	76	.071	126	.056	176	.059
27	.050	77	.070	127	.054	177	.061
28	.049	78	.067	128	.052	178	.059
29	.054	79	.069	129	.050	179	.057
30	.057	80	.067	130	.048	180	.057
31	.055	81	.065	131	.048	181	.057
32	.062	82	.062	132	.053	182	.063
33	.060	83	.062	133	.056	183	.067
34	.060	84	.060	134	.057	184	.067
35	.059	85	.058	135	.055	185	.065
36	.063	86	.056	136	.063	186	.063
37	.061	87	.054	137	.062	187	.062
38	.060	88	.055	138	.059	188	.061
39	.058	89	.052	139	.057	189	.059
40	.058	90	.050	140	.055	190	.059
41	.056	91	.049	141	.053	191	.068
42	.054	92	.047	142	.051	192	.075
43	.053	93	.045	143	.055	193	.074
44	.052	94	.044	144	.053	194	.072
45	.054	95	.049	145	.052	195	.070
46	.054	96	.052	146	.052	196	.073
47	.056	97	.053	147	.060	197	.071
48	.054	98	.052	148	.060		
49	.058	99	.053	149	.058		
50	.056	100	.052	150	.057		

$\hat{V}_{\varepsilon,t}$ produced by VRM ($\beta = 1.4$, $\hat{V}_{\varepsilon,0} = .065$)

when applied to B/J Data

minimum computational effort. The effect of even large discontinuities in the parameters of the underlying process was also insignificant which implies that FRM can overcome all the problems which arise in the MSM operating with a fixed initially nominated estimate of the variance. The effect of a lower limit p_L on the posterior probabilities $p_t^{(k)}$ was investigated and it was found that while it accelerates the response of the method to a genuine discontinuity in the noise variance, it does not affect the on line estimation of the variance when this is relatively stable. As p_L increases the on line estimates become more unstable and a value of $p_L = .0001$ was recommended producing a good balance in the trade off between instability and speed of response.

A special case of the VRM was fully described and recommended using a single parameter β, in order to construct a two point variable range at each point in time. A value of $\beta = 1.4$ was recommended producing a good balance in the trade off between instability and speed of response. Initially VRM operating with $\beta = 1.4$ may be slow to converge to the true value of the variance if the initial crude estimate of the variance is in error by a large factor, say four or greater. It has therefore been suggested to use FRM initially until time $t = 20$ when $\hat{V}_{\varepsilon,20}$ produced by FRM can be used as an initial estimate for VRM. VRM is also very robust to discontinuities in the parameters of the underlying process thus overcoming the problems of the MSM as described in the last section of Chapter 5. Given a genuine discontinuity in the noise variance the VRM with $\beta = 1.4$ responds in a similar way to FRM with $p_L = .0001$. Generally, the recommended forms of FRM and

VRM produce very similar forecasting performance and this was illustrated using 3M Data and B/J Data. The VRM has however a significant computational advantage over FRM, requiring less than one third of the CPU time taken by the latter method and should therefore be preferred.

CHAPTER 7

ESTIMATION OF VARIANCES IN THE STEADY STATE DLM

7.1. Foreword

This chapter examines the problem of on line variance estimation in the single state SSM described in chapter 2.

In section 2.3 it was seen that the SSM assuming constant process variances V_ε, V_μ tends to a limiting form which is equivalent to an exponentially weighted moving average (EWMA) with smoothing constant α such that,

$$\alpha = (-1 + [1 + 4\ r_{\mu,N}]^{-\frac{1}{2}}) \ /\ (2r_{\mu,N}) \tag{7.1.1}$$

where $r_{\mu,N} = V_{\varepsilon,N} \ /\ V_{\mu,N}$

and $V_{\varepsilon,N}$, $V_{\mu,N}$ are initially fixed nominated estimates of the true process variances V_ε, V_μ respectively.

A major advantage of the Bayesian DLM formulation of the SSM over EWMA, is the distributional information available from the Kalman filter, making it possible for decisions to be made based not only on single figure forecasts but probability distributions. Consider for example the distribution of the one step ahead forecast at time t − 1 :

$$(y_t \mid D_{t-1}) \sim N(\hat{y}_t , \hat{Y}_t)$$

From section 2.3 it was seen that in the limit, $\hat{Y}_t$ is independent of the data and depends only on the nominated values $V_{\varepsilon,N}$ and $r_{\mu,N}$:

$$\lim_{t \to \infty} \hat{Y}_t = \hat{Y}_N = V_{\varepsilon,N} \;/\; (1 - A_N) \tag{7.1.2}$$

$$\text{with } A_N = (- 1 + [1 + 4r_{\mu,N}]^{\frac{1}{2}}) \;/\; (2r_{\mu,N}) \tag{7.1.3}$$

The implication is that if $V_{\varepsilon,N}$ and $r_{\mu,N}$ are in some error from the true process parameters V_ε and r_μ (= V_ε / V_μ), then the Kalman filter can produce misleading distributional information even when the MSE is minimised as illustrated by the following example.

Suppose that in trying to model a steady state process we use the SSM with $r_{\mu,N}$ accidentally equal to the true variance ratio r_μ but with $V_{\varepsilon,N}$ overestimating the true noise variance V_ε by a factor of 2 (i.e. $V_{\varepsilon,N} = 2\,V_\varepsilon$). Consider also the limiting value of the MSE as given by (2.3.8) :

$$\lim_{t \to \infty} (\text{MSE})_t = \frac{(2A_N\, V_\varepsilon + V_\mu)}{A_N\,(2 - A_N)} \tag{7.1.4}$$

It follows that the MSE will be optimal since its limiting value depends only on A_N which in turn depends only on $r_{\mu,N}$ (see equation (7.1.3)) and in this example we have assumed that $r_{\mu,N} = r_{\mu}$. At the same time however using (7.1.2) it can be seen that the forecast variance $\hat{Y}_N$ overestimates the true variance by a factor of 2. This has serious cost implications in stock control and planning since if the forecast variance is to be used for calculating confidence intervals on the one step ahead forecast $\hat{y}_t$, then decisions resulting from ,

$$(y_t \mid D_{t-1}) \sim N(\hat{y}_t \, , \, 2V_\varepsilon \; / \; (1 - A_N))$$

are likely to be quite different to those taken, given the correct distribution :

$$(y_t \mid D_{t-1}) \sim N(\hat{y}_t \, , \, V_\varepsilon \; / \; (1 - A_N)$$

It is therefore essential for the SSM to take account of the actual data in order for the distributional information to be meaningful and near optimal. This necessitates on line variance estimation procedures which in addition can be used even when insufficient observations exist for traditional estimation procedures.

Although on line estimation methods are crucial for the successful implementation of single state Bayesian models using the DLM formulation, very little attention has been paid to them in the literature with the exception of Leonard and Harrison [30]. In section 7.2 the concepts of FRM and VRM are applied to the SSM producing on line estimates of V_ε given a fixed nominated estimate of $r_\mu (=V_\varepsilon/V_\mu)$.

In section 7.3 a class I multi process model (see section 1.4 for definition) is proposed producing on line estimates of V_ε and r_μ. Finally in section 7.4 a computationally efficient method is described which produces on line V_ε and r_μ estimates making use of the limiting properties of the SSM. The latter method however is based on the assumption that the true process variances V_ε, V_μ are constant, in contrast to the methods of sections 7.2 and 7.3 which allow these variances to vary through time.

7.2. On line estimation of V_ε in the SSM

In most industrial applications of the SSM the variance ratio r_μ is of the order of 10 or greater (corresponding to EWMA $\alpha \leqslant .27$). Under such circumstances it can be shown (see Appendix K) that the performance of the SSM is insensitive to the choice of our nominated estimate of r_μ and therefore fixing $r_{\mu,N}$ initially, is not likely to be critical to the performance. Hence the FRM and VRM procedures described in the previous chapter for the MSM can easily be adapted to produce on line estimates of V_ε (i.e. $\hat{V}_{\varepsilon,t}$) which can be used by the SSM instead of a fixed $V_{\varepsilon,N}$ nominated value. The modified FRM and VRM can then be viewed as means of on line optimisation of the distributional information such as $\hat{Y}_t$, C_t e.t.c.

7.2.1. FRM in the SSM

Consider the SSM first described in 2.3:

$$\left.\begin{aligned} y_t &= \mu_t + \varepsilon_t & \varepsilon_t &\sim N\,(0,\ V_{\varepsilon,t}) \\ \mu_t &= \mu_{t-1} + \delta\mu_t & \delta\mu_t &\sim N\,(0,\ V_{\mu,t}) \end{aligned}\right\} \qquad (7.2.1.1)$$

where $V_{\varepsilon,t}$ and $V_{\mu,t}$ are the true observation noise and level disturbance variances respectively.

Let the prior information at time t be of the following form:

$$(\mu_{t-1} \mid D_{t-1}) \sim N\,(m_{t-1},\ C_{t-1}) \qquad (7.2.1.2)$$

and $$p\,(V_\varepsilon^{(k)} \mid D_{t-1}) = p_{t-1}^{(k)} \qquad (7.2.1.3)$$

We now postulate that the actual observation at time t, y_t, may come from one of a set of predictive distributions:

$$(y_t \mid V_\varepsilon^{(k)},\ D_{t-1}) \sim N\,(\hat{y}_t\ ,\ \hat{Y}_t^{(k)}) \qquad (7.2.1.4)$$

$$\text{with} \begin{cases} \hat{y}_t = m_{t-1} & (7.2.1.5) \\ \hat{Y}_t^{(k)} = C_{t-1} + V_\mu^{(k)} + V_\varepsilon^{(k)} & (7.2.1.6) \\ \text{for } k = 1, 2, \ldots . K \end{cases}$$

where $V_\varepsilon^{(k)}$ are the values in our V range and $V_\mu^{(k)} = V_\varepsilon^{(k)} / r_{\mu,N}$.

The variance ratio $r_{\mu,N}$ is nominated initially as our best estimate of the true variance ratio r_μ (= $V_{\varepsilon,t}/V_{\mu,t}$) amd remains fixed. The effect of $r_{\mu,N}$ on the performance of the SSM has been shown (see Appendix K) to be insignificant and hence errors in $r_{\mu,N}$ will not be critical to the on line estimation of the noise variance.

As soon as y_t becomes available we can calculate the likelihoods of it having come from each one of the predictive distributions postulated in (7.2.1.4) :

$$L\ (y_t \mid \hat{Y}_t^{(k)}, D_{t-1}) \propto \left[\hat{Y}_t^{(k)} \right]^{-\frac{1}{2}} \exp \{-\tfrac{1}{2}\, e_t^2 / \hat{Y}_t^{(k)} \} \qquad (7.2.1.7)$$

where e_t = forecast error at time t = $y_t - \hat{y}_t$

Knowledge of $V_\varepsilon^{(k)}$ however, is equivalent to $\hat{Y}_t^{(k)}$ and therefore,

$$L\ (y_t \mid V_\varepsilon^{(k)}, D_{t-1}) = L\ (y_t \mid \hat{Y}_t^{(k)}, D_{t-1}) \qquad (7.2.1.8)$$

and this result enables us to find the probability distribution of y_t conditional on $V_\varepsilon^{(k)}$ for k = 1, 2, ... K as follows:

$$p(y_t \mid V_\varepsilon^{(k)}, D_{t-1}) = (1/k_1) \quad L(y_t \mid V_\varepsilon^{(k)}, D_{t-1}) \tag{7.2.1.9}$$

$$\text{where } k_1 = \sum_{k=1}^{K} L(y_t \mid V_\varepsilon^{(k)}, D_{t-1}) \tag{7.2.1.10}$$

Finally posterior probabilities $p_t^{(k)}$ for the $V_\varepsilon^{(k)}$ values in our V range can be calculated as follows:

$$p(V_\varepsilon^{(k)} \mid y_t , D_{t-1}) = k_2 \; p \; (y_t \mid V_\varepsilon^{(k)}, D_{t-1}) \; p \; (V_\varepsilon^{(k)} \mid D_{t-1})$$

or using established notation,

$$p_t^{(k)} = k_2 \cdot p(y_t \mid V_\varepsilon^{(k)}, D_{t-1}) \cdot p_{t-1}^{(k)} \tag{7.2.1.11}$$

where the second and third terms of this expression are known from (7.2.1.9) and (7.2.1.3) respectively and k_2 is a constant :

$$k_2 = 1/p \; (y_t \mid D_{t-1}) = \left[\sum_k p(y_t \mid V_\varepsilon^{(k)}, D_{t-1}) \cdot p_{t-1}^{(k)} \right]^{-1}$$

Hence our best line estimate of the noise variance can now be calculated as :

$$\hat{V}_{\varepsilon,t} = \sum_k p_t^{(k)} V_\varepsilon^{(k)} \tag{7.2.1.12}$$

and using $\hat{V}_{\varepsilon,t}$ in the Kalman filter equations we can update the prior information given by (7.2.1.2) as follows:

$$\left.\begin{aligned}
R_t &= C_{t-1} + \hat{V}_{\varepsilon,t} / r_{\mu,N} \\
\hat{Y}_t &= R_t + \hat{V}_{\varepsilon,t} \\
A_t &= R_t / \hat{Y}_t \\
m_t &= m_{t-1} + A_t \; e_t \\
C_t &= A_t \; \hat{V}_{\varepsilon,t}
\end{aligned}\right\} \tag{7.2.1.13}$$

We now have prior information for time period t+1 as follows:

$$(\mu_t \mid D_t) \sim N(m_t \, , \, C_t) \tag{7.2.1.14}$$

and $p(V_\varepsilon^{(k)} \mid D_t) = p_t^{(k)}$ (7.2.1.15)

Note that (7.2.1.14) and (7.2.1.15) are in exactly the same form as (7.2.1.2) and (7.2.1.3) respectively except that the time subscript is now t instead of t - 1. A whole cycle has therefore been completed and the updating system described in equations (7.2.1.4) to (7.2.1.13) can be repeated with each new observation.

A computer program listing for this procedure is given in Appendix H.

7.2.2. Numerical illustration

The experimentation performed here is very similar to that reported earlier in chapter 6 for FRM in relation to MSM. Most of the conclusions and recommendations also are valid here and therefore to avoid repetition we will simply illustrate the method using C6 Data and paying attention mainly to areas where the FRM/SSM (we will from now on refer to FRM in relation to MSM and SSM by FRM/MSM and FRM/SSM respectively) is in one way or another different to FRM/MSM.

The following are for example also true for FRM/SSM:

(i) The choice of c and K in the construction of our V range is not critical to the performance of MSM/SSM provided $K \geqslant 11$ and $c \leqslant 1.5$ and the particular case of $K = 11$ and $c = 1.5$ can be recommended on grounds of minimum computational effort.

(ii) The effect of our initial crude estimate V_o dies away very quickly and $\hat{V}_{\varepsilon,t}$ is then wholly determined by the actual data.

Let us now examine the effect of our nominated estimate $r_{\mu,N}$ which in FRM/SSM is initially fixed. Consider C6 Data where the true variance ratio is $r_{\mu} = 25$ and suppose we apply FRM/SSM with $r_{\mu,N} = 10$, 25 and 100. The on line estimates $\hat{V}_{\varepsilon,t}$ produced are tabulated in table 7.1 for $t = 1, 2, \ldots 100$. It can be seen that the different choices of

Table 7.1

t	$r_{\mu,N} = 10$	25	100
1	.0195	.0195	.0196
2	.0148	.0149	.0150
3	.0101	.0102	.0103
4	.0068	.0069	.0069
5	.0048	.0048	.0049
6	.0035	.0035	.0035
7	.0033	.0034	.0034
8	.0028	.0029	.0029
9	.0030	.0031	.0031
10	.0026	.0027	.0027
11	.0023	.0024	.0024
12	.0021	.0021	.0021
13	.0021	.0021	.0021
14	.0021	.0021	.0022
15	.0027	.0028	.0028
16	.0025	.0025	.0026
17	.0024	.0024	.0025
18	.0027	.0028	.0029
19	.0025	.0026	.0027
20	.0026	.0028	.0029
21	.0026	.0027	.0028
22	.0024	.0025	.0026
23	.0023	.0024	.0024
24	.0023	.0024	.0026
25	.0022	.0023	.0025
26	.0021	.0022	.0024
27	.0020	.0021	.0023
28	.0019	.0020	.0022
29	.0019	.0020	.0021
30	.0019	.0020	.0022
31	.0019	.0020	.0021
32	.0019	.0019	.0021
33	.0018	.0019	.0020
34	.0017	.0018	.0019
35	.0018	.0019	.0020
36	.0017	.0018	.0019
37	.0017	.0018	.0020
38	.0022	.0023	.0025
39	.0022	.0023	.0024
40	.0027	.0029	.0030
41	.0027	.0029	.0030
42	.0027	.0028	.0030
43	.0027	.0029	.0030
44	.0026	.0028	.0029
45	.0026	.0027	.0029
46	.0025	.0027	.0028
47	.0025	.0027	.0028
48	.0025	.0026	.0027
49	.0024	.0025	.0027
50	.0025	.0027	.0029
51	.0025	.0026	.0028
52	.0025	.0027	.0028
53	.0026	.0028	.0029
54	.0026	.0028	.0030
55	.0025	.0028	.0031
56	.0025	.0028	.0030
57	.0025	.0027	.0029
58	.0025	.0027	.0029
59	.0025	.0027	.0030
60	.0025	.0027	.0030
61	.0025	.0027	.0030
62	.0024	.0027	.0030
63	.0024	.0027	.0030
64	.0025	.0027	.0030
65	.0024	.0026	.0030
66	.0024	.0026	.0029
67	.0024	.0026	.0029
68	.0024	.0026	.0030
69	.0024	.0026	.0029
70	.0026	.0028	.0031
71	.0025	.0028	.0031
72	.0025	.0027	.0030
73	.0025	.0027	.0030
74	.0024	.0026	.0030
75	.0024	.0026	.0029
76	.0024	.0026	.0029
77	.0023	.0026	.0029
78	.0024	.0026	.0029
79	.0023	.0025	.0029
80	.0023	.0025	.0029
81	.0024	.0026	.0029
82	.0024	.0026	.0029
83	.0024	.0026	.0029
84	.0023	.0025	.0029
85	.0023	.0026	.0029
86	.0023	.0025	.0029
87	.0023	.0025	.0029
88	.0023	.0025	.0029
89	.0024	.0026	.0029
90	.0023	.0025	.0029
91	.0023	.0025	.0028
92	.0023	.0025	.0028
93	.0023	.0025	.0028
94	.0023	.0026	.0029
95	.0023	.0025	.0029
96	.0023	.0025	.0029
97	.0023	.0025	.0028
98	.0023	.0024	.0028
99	.0022	.0024	.0028
100	.0022	.0024	.0028

$\hat{V}_{E,t}$ produced by FRM/SSM when applied to CG Data, using $r_{\mu,N}$ = 10, 25 and 100. The V range used is that given by (6.2.2.1) with V_0 = .0100

$r_{\mu,N}$ result in relatively similar on line estimates $\hat{V}_{\varepsilon,t}$ with $r_{\mu,N} = r_{\mu} = 25$ converging and fluctuating around the true value of $V_{\varepsilon} = .0025$ while $r_{\mu,N} = 10$ and $r_{\mu,N} = 100$ converge to levels which slightly underestimate and overestimate the true noise variance, respectively. This response is consistent with the theoretical results of Appendix K, where from figure K.1 for example it can be seen that if r_{μ} is overestimated ($k_1 > 1$) then the maximum likelihood estimate of V_{ε} is also overestimated ($k_2 > 1$) with the opposite results when r_{μ} is underestimated. Appendix K also concludes that the performance of the method is not significantly affected by the choise of $r_{\mu,N}$ ($V_{\varepsilon,N}$ being much more significant than $r_{\mu,N}$ provided r_{μ} is larger than approximately ten), and the results of table 7.1 clearly illustrate this point. Finally it was shown in Appendix K that as r_{μ} increases ($r_{\mu} = 25$ in C6 Data) the choise of $r_{\mu,N}$ becomes less significant and this implies that for larger r_{μ}, the FRM/SSM over/underestimation becomes even less serious.

Next we illustrate the response of FRM/SSM to a discontinuity in V_{ε}, where slightly different recommendations about a lower probability limit p_L, is made than that of FRM/MSM in chapter 6. Recall that p_L was introduced to enable FRM/MSM to respond faster to step changes in the noise variance V_{ε}. Consider C6 Data generated with:

$$V_{\varepsilon} = \begin{cases} .0025 & t \leq 100 \\ .0050 & t > 100 \end{cases},$$

The on line estimates $\hat{V}_{\varepsilon,t}$ for t = 2, 4, ... 200 corresponding to different choices of p_L, and shown in table 7.2 which is directly comparable with table 6.13 for FRM/MSM. Comparing these two tables it becomes clear that FRM/SSM is more sensitive to p_L than FRM/MSM. This is because the SSM does not model discontinuities and therefore every relatively large forecast error (say over 2σ) is interpreted by FRM as a possible signal of increased variability, while FRM/MSM is not affected by such errors. It is therefore recommended to use a smaller value of p_L in FRM/SSM and $p_L = .00001$ has been found to produce a good balance in the trade off between stability and speed of response. For values of $p_L > 0$ up to the recommended value, the conclusion drawn for FRM/MSM is still valid, that is the effect of p_L on $\hat{V}_{\varepsilon,t}$ is insignificant when V_ε is relatively stable but at points of V_ε discontinuity p_L accelerates the convergence of the on line estimates to the new V_ε level.

Table 7.2

t	$\rho_L = 0$	$\rho_L = 10^{-6}$	$\rho_L = 10^{-5}$	$\rho_L = 10^{-4}$	$\rho_L = 10^{-3}$
2	.0149	.0149	.0149	.0149	.0149
4	.0069	.0069	.0069	.0069	.0070
6	.0035	.0035	.0035	.0035	.0036
8	.0029	.0029	.0029	.0029	.0030
10	.0027	.0027	.0027	.0027	.0029
12	.0021	.0021	.0021	.0021	.0022
14	.0021	.0021	.0021	.0022	.0023
16	.0025	.0025	.0026	.0026	.0029
18	.0028	.0028	.0028	.0029	.0033
20	.0028	.0028	.0028	.0028	.0031
22	.0025	.0025	.0025	.0025	.0026
24	.0024	.0024	.0025	.0025	.0026
26	.0022	.0022	.0022	.0022	.0023
28	.0020	.0020	.0020	.0021	.0021
30	.0020	.0020	.0020	.0020	.0022
32	.0019	.0019	.0019	.0020	.0021
34	.0018	.0018	.0018	.0018	.0019
36	.0018	.0018	.0018	.0018	.0019
38	.0023	.0023	.0024	.0027	.0058
40	.0029	.0029	.0030	.0045	.0116
42	.0028	.0028	.0029	.0033	.0064
44	.0028	.0028	.0028	.0030	.0042
46	.0027	.0027	.0027	.0027	.0032
48	.0026	.0026	.0026	.0026	.0029
50	.0027	.0027	.0027	.0027	.0032
52	.0027	.0027	.0027	.0027	.0029
54	.0028	.0028	.0028	.0028	.0032
56	.0028	.0028	.0028	.0028	.0030
58	.0027	.0027	.0027	.0027	.0029
60	.0027	.0027	.0027	.0027	.0030
62	.0027	.0027	.0027	.0026	.0028
64	.0027	.0027	.0027	.0027	.0028
66	.0026	.0026	.0026	.0026	.0026
68	.0026	.0026	.0026	.0026	.0028
70	.0028	.0028	.0028	.0029	.0038
72	.0027	.0027	.0027	.0027	.0029
74	.0026	.0026	.0026	.0026	.0027
76	.0026	.0026	.0026	.0026	.0026
78	.0026	.0026	.0026	.0026	.0027
80	.0025	.0025	.0025	.0025	.0026
82	.0026	.0026	.0026	.0026	.0028
84	.0025	.0025	.0025	.0025	.0026
86	.0025	.0025	.0025	.0025	.0026
88	.0025	.0025	.0025	.0025	.0026
90	.0025	.0025	.0025	.0025	.0026
92	.0025	.0025	.0025	.0025	.0025
94	.0026	.0026	.0026	.0026	.0028
96	.0025	.0025	.0025	.0025	.0026
98	.0024	.0024	.0024	.0024	.0025
100	.0024	.0024	.0024	.0024	.0024

$\hat{V}_{\varepsilon,t}$ produced by FRM|SSM when applied to C6 Data generated with $V_\varepsilon = \begin{cases} .0025 & \text{for } t \le 100 \\ .0050 & \text{for } t > 100 \end{cases}$. The V range used is that given by (6.2.2.1) with $V_0 = .0100$.

$r_{M,N} = 25$

Table 7.2 cont.

102	.0027	.0027	.0027	.0030	.0061
104	.0028	.0029	.0029	.0035	.0084
106	.0028	.0028	.0028	.0030	.0049
108	.0028	.0028	.0028	.0029	.0036
110	.0028	.0028	.0028	.0029	.0037
112	.0028	.0028	.0028	.0029	.0034
114	.0028	.0028	.0028	.0028	.0032
116	.0029	.0029	.0029	.0029	.0036
118	.0030	.0030	.0030	.0036	.0071
120	.0031	.0031	.0033	.0047	.0085
122	.0032	.0034	.0044	.0081	.0115
124	.0035	.0041	.0067	.0103	.0118
126	.0034	.0036	.0045	.0068	.0086
128	.0034	.0035	.0040	.0058	.0075
130	.0040	.0050	.0079	.0096	.0106
132	.0039	.0045	.0066	.0083	.0090
134	.0040	.0045	.0064	.0079	.0084
136	.0040	.0044	.0061	.0075	.0079
138	.0041	.0046	.0064	.0075	.0079
140	.0041	.0046	.0061	.0072	.0075
142	.0041	.0043	.0053	.0063	.0066
144	.0040	.0042	.0049	.0059	.0062
146	.0040	.0041	.0047	.0056	.0059
148	.0040	.0041	.0045	.0053	.0056
150	.0040	.0041	.0044	.0051	.0054
152	.0039	.0039	.0041	.0046	.0047
154	.0040	.0041	.0044	.0050	.0058
156	.0040	.0041	.0043	.0049	.0054
158	.0041	.0042	.0045	.0052	.0059
160	.0041	.0041	.0043	.0048	.0053
162	.0041	.0041	.0043	.0048	.0052
164	.0042	.0043	.0046	.0052	.0058
166	.0043	.0043	.0046	.0052	.0058
168	.0042	.0042	.0044	.0048	.0052
170	.0042	.0042	.0043	.0046	.0049
172	.0041	.0041	.0042	.0044	.0047
174	.0041	.0041	.0041	.0043	.0045
176	.0042	.0042	.0043	.0046	.0052
178	.0044	.0044	.0049	.0057	.0078
180	.0044	.0044	.0048	.0055	.0067
182	.0044	.0044	.0047	.0053	.0060
184	.0044	.0044	.0047	.0052	.0057
186	.0044	.0044	.0046	.0051	.0055
188	.0043	.0044	.0045	.0049	.0051
190	.0044	.0045	.0048	.0054	.0062
192	.0044	.0045	.0048	.0055	.0061
194	.0044	.0045	.0049	.0056	.0063
196	.0044	.0045	.0050	.0058	.0066
198	.0044	.0045	.0050	.0057	.0062
200	.0044	.0045	.0050	.0057	.0062

7.2.3. VRM in the SSM

Suppose that posterior to y_{t-1} and just before y_t is known, we have a best on line estimate $\hat{V}_{\varepsilon,t-1}$ of the true variance $V_{\varepsilon,t-1}$, i.e.

$$\hat{V}_{\varepsilon,t-1} = E\,(V_{\varepsilon,t-1} \mid D_{t-1}) \qquad (7.2.3.1)$$

This estimate is now used in the Kalman filter equations (just like in equations (7.2.1.13) for the FRM) to obtain the posterior distribution of the true process level μ_{t-1} :

$$(\mu_{t-1} \mid D_{t-1}) \sim N(m_{t-1}\ ,\ C_{t-1}) \qquad (7.2.3.2)$$

We can now construct our time t variable range $(V^{(1)}_{\varepsilon,t}\ ,\ V^{(2)}_{\varepsilon,t})$ using constants c_1 , c_2 and associated prior probabilities $p^{(1)}$ and $p^{(2)}$ in the same way described earlier in section 6.3 i.e.

$$\left.\begin{aligned} V^{(1)}_{\varepsilon,t} &= c_1\ \hat{V}_{\varepsilon,t-1} \\ V^{(2)}_{\varepsilon,t} &= c_2\ \hat{V}_{\varepsilon,t-1} \end{aligned}\right\} \qquad (7.2.3.3)$$

$$\left.\begin{aligned} \text{and } p^{(1)} &= p\,(V^{(1)}_{\varepsilon,t} \mid D_{t-1}) = \frac{c_2-1}{c_2-c_1} \\ p^{(2)} &= p\,(V^{(2)}_{\varepsilon,t} \mid D_{t-1}) = \frac{1-c_1}{c_2-c_1} \end{aligned}\right\} \qquad (7.2.3.4)$$

such that the weighted average of $V_{\varepsilon,t}^{(1)}$ and $V_{\varepsilon,t}^{(2)}$ prior to observing y_t is equal to our latest on line estimate of the variance, $\hat{V}_{\varepsilon,t-1}$:

$$\sum_{k=1}^{2} p^{(k)} V_{\varepsilon,t}^{(k)} = \hat{V}_{\varepsilon,t-1} \qquad (7.2.3.5)$$

As soon as y_t becomes available it is viewed by VRM as having come from one of the following two predictive distributions:

$$(y_t \mid D_{t-1}) \sim N(\hat{y}_t , \hat{Y}_t^{(k)}) \qquad \text{for } k = 1, 2 \qquad (7.2.3.6)$$

where $\hat{y}_t$ and $\hat{Y}_t^{(k)}$ are as given by (7.2.1.5) and (7.2.1.6) for FRM earlier In particular $\hat{Y}_t^{(k)}$ for k = 1, 2 corresponds to $V_{\varepsilon,t}^{(1)}$ and $V_{\varepsilon,t}^{(2)}$

We can then calculate likelihoods $L_t^{(k)}$:

$$L_t^{(k)} = L\,(y_t \mid V_{\varepsilon,t}^{(k)} , D_{t-1}) \propto$$

$$\propto \quad [\hat{Y}_t^{(k)}]^{-\frac{1}{2}} \quad \exp.\ \{-\tfrac{1}{2}\ e_t^2 / \hat{Y}_t^{(k)}\} \qquad (7.2.3.7)$$

If we now let $p_t^{(k)}$ be the posterior probabilities for $V_{\varepsilon,t}^{(k)}$, k = 1, 2 then we have :

$$p_t^{(k)} = p\,(V_{\varepsilon,t}^{(k)} \mid y_t\,,\, D_{t-1}) = p(y_t \mid V_{\varepsilon,t}^{(k)}\,,\, D_{t-1}) \,.\, p\,(V_{\varepsilon,t}^{(k)} \mid D_{t-1})$$

and using (7.2.3.7) and (7.2.3.4) we get:

$$\left. \begin{array}{l} p_t^{(k)} = [\,\Sigma\, L_t^{(k)} \quad p^{(k)}]^{-1} \quad .\, L_t^{(k)} \quad .\, p^{(k)} \quad \} \\ \text{for } k = 1,\, 2. \end{array} \right\} \quad (7.2.3.8)$$

The posterior estimateof the true observation noise variance at time t can now be calculated as the weighted average of the two postulated variances :

$$\hat{V}_{\varepsilon,t} = E\,(V_{\varepsilon,t} \mid D_t) = \sum_{k=1}^{2} p_t^{(k)} \quad V_{\varepsilon,t}^{(k)} \qquad (7.2.3.9).$$

At this point we have completed a whole cycle of the updating process since the estimate given by (7.2.3.9) can now be used in the Kalman filter to obtain the posterior distribution of μ_t:

$$(\mu_t \mid D_t) \sim N\,(m_t\,,\, C_t) \qquad (7.2.3.10)$$

which is in the same form as (7.2.3.2.) except that the time index is now t instead of t - 1. As soon as y_{t+1} becomes available the procedure described in equations (7.2.3.3) to (7.2.3.10) can be repeated and similarly for every new observation.

A computer program listing for this procedure is given in Appendix H.

7.2.4. Numerical Illustration

The important factor here as with VRM/MSM (VRM/MSM and VRM/SSM will be used to distinguish VRM in relation to MSM and SSM respectively) is the specification of the constants c_1 , c_2 which define the variable range $(V^{(1)}_{\varepsilon,t}$, $V^{(2)}_{\varepsilon,t})$ at each point in time. The best and most convenient way of specifying c_1 ,c_2 was again found to be in terms of a single parameter β, i.e.

$$c_1 = {}^1/\beta \text{ and } c_2 = \beta$$

Considerable experimentation using different specifications for c_1 , c_2 as well as more than two values in our variable range $(V^{(k)}_{\varepsilon,t}$; $k = 1, 2, \ldots K)$ at time t, did not lead to significantly better results and is therefore not reported.

For the same reason that in FRM/SSM a smaller value for p_L was recommended than in FRM/MSM, we now expect that a smaller value of β will be appropriate in VRM/SSM than $\beta = 1.4$ which was recommended in VRM/MSM. The on line estimates $\hat{V}_{\varepsilon,t}$ produced by VRM/SSM using a large range of β, on C6 Data with a discontinuity in V_ε at time $t = 100$, i.e.

$$V_\varepsilon = \begin{cases} .0025 & t \leq 100 \\ .0050 & t > 100 \end{cases}$$

are reported in table 7.3 for $t = 2, 4, \ldots .200$.

Table 7.3

t	$\beta=1.2$	$\beta=1.3$	$\beta=1.4$	$\beta=1.5$	$\beta=1.6$
2	.0050	.0049	.0049	.0048	.0048
4	.0049	.0047	.0045	.0043	.0041
6	.0047	.0045	.0042	.0038	.0035
8	.0047	.0044	.0040	.0037	.0034
10	.0046	.0043	.0039	.0035	.0032
12	.0045	.0040	.0035	.0031	.0027
14	.0044	.0040	.0035	.0031	.0027
16	.0045	.0040	.0036	.0033	.0031
18	.0045	.0040	.0037	.0035	.0034
20	.0044	.0039	.0036	.0033	.0032
22	.0043	.0038	.0033	.0030	.0028
24	.0042	.0037	.0032	.0029	.0027
26	.0041	.0035	.0030	.0026	.0024
28	.0040	.0033	.0027	.0023	.0020
30	.0040	.0032	.0027	.0023	.0020
32	.0039	.0031	.0025	.0021	.0019
34	.0038	.0030	.0023	.0019	.0016
36	.0037	.0029	.0023	.0019	.0017
38	.0039	.0033	.0029	.0027	.0026
40	.0041	.0036	.0035	.0036	.0037
42	.0040	.0036	.0034	.0034	.0035
44	.0040	.0035	.0033	.0032	.0032
46	.0039	.0033	.0030	.0029	.0027
48	.0038	.0032	.0029	.0027	.0025
50	.0038	.0033	.0030	.0029	.0029
52	.0038	.0032	.0029	.0028	.0027
54	.0038	.0033	.0031	.0031	.0031
56	.0038	.0033	.0031	.0030	.0030
58	.0037	.0032	.0029	.0028	.0028
60	.0037	.0032	.0030	.0029	.0029
62	.0036	.0030	.0028	.0026	.0026
64	.0036	.0030	.0028	.0027	.0026
66	.0035	.0029	.0026	.0024	.0022
68	.0035	.0029	.0027	.0026	.0025
70	.0036	.0032	.0031	.0032	.0034
72	.0035	.0030	.0029	.0029	.0029
74	.0034	.0029	.0027	.0026	.0025
76	.0034	.0028	.0025	.0023	.0022
78	.0033	.0027	.0025	.0023	.0022
80	.0033	.0027	.0024	.0022	.0021
82	.0033	.0028	.0026	.0025	.0026
84	.0032	.0026	.0024	.0023	.0022
86	.0032	.0026	.0024	.0023	.0022
88	.0032	.0026	.0023	.0022	.0022
90	.0032	.0026	.0024	.0023	.0023
92	.0031	.0025	.0023	.0021	.0021
94	.0032	.0026	.0025	.0025	.0025
96	.0031	.0026	.0024	.0024	.0024
98	.0030	.0025	.0022	.0021	.0021
100	.0030	.0024	.0021	.0020	.0019

$\hat{V}_{\varepsilon,t}$ produced by VRM|SSM when applied to C6 Data generated with $V_\varepsilon = \begin{cases} .0025 & t \le 100 \\ .0050 & t > 100 \end{cases}$. $\hat{V}_{\varepsilon,0} = .0050$ and $r_{\mu,N} = 25$

Table 7.3 cont.

102	.0032	.0028	.0028	.0028	.0029
104	.0034	.0032	.0033	.0036	.0039
106	.0033	.0031	.0032	.0033	.0035
108	.0033	.0030	.0030	.0032	.0032
110	.0033	.0031	.0032	.0033	.0034
112	.0033	.0030	.0031	.0032	.0033
114	.0032	.0030	.0030	.0030	.0030
116	.0033	.0031	.0032	.0033	.0034
118	.0036	.0037	.0041	.0047	.0052
120	.0038	.0041	.0048	.0056	.0063
122	.0041	.0047	.0057	.0068	.0079
124	.0043	.0052	.0064	.0077	.0089
126	.0042	.0049	.0059	.0067	.0074
128	.0042	.0048	.0056	.0063	.0068
130	.0046	.0057	.0069	.0081	.0091
132	.0046	.0055	.0066	.0075	.0081
134	.0046	.0055	.0065	.0072	.0076
136	.0046	.0055	.0063	.0069	.0072
138	.0047	.0056	.0064	.0070	.0072
140	.0047	.0055	.0063	.0067	.0068
142	.0046	.0052	.0057	.0059	.0057
144	.0045	.0051	.0055	.0055	.0052
146	.0045	.0050	.0053	.0052	.0049
148	.0044	.0049	.0051	.0049	.0046
150	.0044	.0048	.0049	.0047	.0044
152	.0043	.0046	.0045	.0042	.0037
154	.0044	.0048	.0049	.0048	.0046
156	.0044	.0047	.0048	.0046	.0044
158	.0045	.0049	.0051	.0051	.0051
160	.0044	.0047	.0047	.0046	.0045
162	.0044	.0047	.0047	.0045	.0044
164	.0046	.0050	.0052	.0052	.0053
166	.0046	.0050	.0052	.0053	.0054
168	.0044	.0047	.0047	.0046	.0045
170	.0044	.0046	.0045	.0043	.0041
172	.0043	.0044	.0042	.0039	.0037
174	.0042	.0043	.0040	.0037	.0034
176	.0044	.0046	.0045	.0045	.0045
178	.0048	.0054	.0059	.0064	.0071
180	.0048	.0053	.0058	.0062	.0068
182	.0047	.0052	.0055	.0058	.0061
184	.0047	.0051	.0053	.0055	.0057
186	.0046	.0049	.0050	.0051	.0052
188	.0045	.0047	.0047	.0046	.0045
190	.0048	.0053	.0056	.0058	.0061
192	.0048	.0053	.0057	.0060	.0062
194	.0049	.0055	.0059	.0063	.0066
196	.0051	.0058	.0063	.0067	.0072
198	.0050	.0056	.0060	.0063	.0065
[illegible]	.0051	.0056	.0060	.0063	.0065

Note that for $t \leq 100$ and $t > 100$ table 7.3 is directly comparable to tables 6.16 and 6.18 respectively. Comparing tables 7.3 and 6.16 it can be seen that for the same value of β, VRM/SSM produces $\hat{V}_{\varepsilon,t}$ estimates which are more unstable than those produced by VRM/MSM. Again this is a result of the fact that the MSM models discontinuities explicitely and it is therefore much more robust to isolated large forecasting errors. As a result of this sensitivity however, $\hat{V}_{\varepsilon,t}$ produced by VRM/SSM responds faster to discontinuities in V_ε as can be seen by comparing tables 7.3 and 6.18.

From a large number of similar experiments, using realisations of the C6 Data type, it has been established that a value of $\beta = 1.3$ produces a consistently good response and can therefore be recommended as a trade off value in the conflicting requirements between stability and speed of response. Note that the recommended value of β here is smaller than that ($\beta = 1.4$) in VRM/MSM since it was shown that the $\hat{V}_{\varepsilon,t}$ estimates produced by VRM/SSM are more sensitive to β than those produced by the former system.

Finally, the problem of initial convergence given a crude initial estimate $\hat{V}_{\varepsilon,o}$ can be solved in the same way described in section 6.4.3. earlier for VRM/MSM. That is FRM/SSM can be used initially because of its faster initial convergence and the on line estimate produced at some arbitrarily early point in time (say $t = 20$) can then be used as an initial estimate for VRM.

7.3 On line estimation of V_ε and r_μ in the SSM

In Appendix K it was shown that when the true variance ratio $r_\mu(=V_\varepsilon/V_\mu)$ of a steady state process is large then the choice of a fixed initially nominated ratio $r_{\mu,N}$ does not significantly affect the performance of the SSM even when it is in error by a large factor. In this kind of situation only the noise variance is significant and the methods of the previous section have proved satisfactory.

For small r_μ, however, say less than 10, the significance of the choice of $r_{\mu,N}$ increases and better forecasting performance can be achieved by estimating both V_ε and V_μ on line simultaneously. Such a joint estimation method has been designed along the lines of the class I multi state models proposed by H/S and briefly defined in chapter 1. The present method will therefore be referred to from now on as CIM (class I method).

In CIM we use a discrete joint probability distribution for the noise variance V_ε and variance ratio r_μ. Consider again the SSM as specified earlier by (7.2.1.1) :

$$\left.\begin{aligned} y_t &= \mu_t + \varepsilon_t & \varepsilon_t &\sim N(0,V_{\varepsilon,t}) \\ \mu_t &= \mu_{t-1} + \delta\mu_t & \delta\mu_t &\sim N(0,\ V_{\mu,t}) \end{aligned}\right\} \qquad (7.3.1)$$

Let $V_\varepsilon^{(k)}$ for $k = 1, = \ldots. K$ and $r_\mu^{(q)}$ for $q = 1, 2\ldots. Q$ be two sets of values which are thought to cover the likely value of $V_{\varepsilon,t}$ and $r_{\mu,t} = V_{\varepsilon,t}/V_{\mu,t}$ respectively for all times t. We then assume that the correct model (for the process under observation) and its parameters can be adequately described by a linear combination of a set of discrete alternative steady state models which we will denote by $M^{(kq)}$ with associated prior probabilities $p_o^{(kq)}$. The latter will be updated with every new observation so that at time t, $p_t^{(kq)}$ reflects the posterior probabilit of the model $M^{(kq)}$ being the correct one, for the process we are observing.

Model $M^{(kq)}$ is characterised by a noise variance equal to $V_\varepsilon^{(k)}$ and a level disturbance variance equal to $V_\varepsilon^{(k)}/r_\mu^{(q)}$. These variances are fixed within each model and therefore in terms of our established notation (see section 2.3) for a SSM with fixed initially nominated variances, model $M^{(kq)}$ is an SSM operating with $V_{\varepsilon,N} = V_\varepsilon^{(k)}$ and $r_{\mu,N} = r_\mu^{(q)}$ for $k = 1, = \ldots. K$ and $q = 1, = \ldots. Q$.

Before the updating procedure is described we will briefly comment on the basic difference between FRM and CIM. In FRM we have K values $V_\varepsilon^{(k)}$ and associated probabilities which are updated with every new observation to give our best on line estimate posterior to time t-1, $\hat{V}_{\varepsilon,t-1}$, as the weighted average of $V_\varepsilon^{(k)}$. This estimate is then used in the Kalman filter leading to a single distribution for the process level at time t-1:

$$(\mu_{t-1} \mid D_{t-1}) \sim N(m_{t-1}, C_{t-1})$$

It is of course possible to attempt joint estimation of V_ε and r_μ by a straight forward extension of FRM. That is by specifying

$$V_\varepsilon^{(k)} \quad \text{and } r_\mu^{(q)} \qquad (k = 1, 2, \ldots .K, q = 1, =, \ldots ..Q)$$

and obtaining best estimates for V_ε and r_μ. These estimates would then be used in the Kalman filter in order to arrive at a single distribution for μ_{t-1} as given above. However, considerable experimentation has shown that as a result of the insensitivity of the likelihood function to the variance ratio (see Appendix K) it is not possible to obtain good estimates of V_ε and r_μ simultaneously, in this way (i.e. using an extended form of FRM or VRM). It is therefore necessary to employ separate models $M^{(kq)}$ as used by CIM, leading to KQ (instead of a single) distributions for μ_{t-1} conditional on each model $M^{(kq)}$, as will be given by (7.3.2) below.

Let the prior information at time t be of the following form for each model $M^{(kq)}$:

$$(\mu_{t-1} \mid M^{(kq)}, D_{t-1}) \sim N\,(m^{(kq)}, C_{t-1}^{(kq)}) \qquad (7.3.2)$$

and $$p(V_\varepsilon^{(k)}, r_\mu^{(q)} \mid M^{(kq)}, D_{t-1}) = p_{t-1}^{(kq)} \qquad (7.3.3)$$

The predictive distribution of $M^{(kq)}$ is then :

$$(y_t \mid M^{(kq)}, D_{t-1}) \mathrel{\dot{\sim}} N(\hat{y}_t^{(kq)}, \hat{Y}_t^{(kq)}) \tag{7.3.4}$$

$$\left.\begin{aligned} &\text{with} \quad \hat{y}_t^{(kq)} = m_{t-1}^{(kq)} \\ &\text{and} \quad \hat{Y}_t^{(kq)} = C_{t-1}^{(kq)} + V_\varepsilon^{(k)} (1/r_\mu^{(q)} + 1) \end{aligned}\right\} \tag{7.3.5}$$

As soon as y_t becomes available we can calculate the posterior distribution of the true process level as viewed by each model :

$$(\mu_t \mid M^{(kq)}, D_{t-1}) \sim N(m_t^{(kq)}, C_t^{(kq)}) \tag{7.3.6}$$

$$\left.\begin{aligned} \text{where} \quad m^{(kq)} &= m_{t-1}^{(kq)} + A_t^{(kq)} e_t^{(kq)} \\ C_t^{(kq)} &= A_t^{(kq)} V_\varepsilon^{(k)} \\ e_t^{(kq)} &= y_t - \hat{y}_t^{(kq)} \\ A_t^{(kq)} &= R_t^{(kq)} / \hat{Y}_t^{(kq)} \\ R_t^{(kq)} &= C_{t-1}^{(kq)} + V_\varepsilon^{(k)} / r_\mu^{(q)} \\ Y_t^{(kq)} &= R_t^{(kq)} + V_\varepsilon^{(k)} \end{aligned}\right\} \tag{7.3.7}$$

We can also calculate likelihoods of the actual observation y_t conditional on each model $M^{(kq)}$ by recalling that knowledge of $M^{(kq)}$ is equivalent to $V_\varepsilon^{(k)}$, $r_\mu^{(q)}$ and therefore to $\hat{Y}_t^{(kq)}$ also :

$$\left.\begin{aligned} & L(y_t \mid M^{(kq)}, D_{t-1}) = L(y_t \mid V_\varepsilon^{(k)}, r_\mu^{(q)}, D_{t-1}) \\ & = L(y_t \mid \hat{Y}_t^{(kq)}, D_{t-1}) \\ & \propto [\hat{Y}_t^{(kq)}]^{-\frac{1}{2}} \quad \exp\{-\tfrac{1}{2}[e_t^{(kq)}]^2 / \hat{Y}_t^{(kq)}\} \end{aligned}\right\} \quad (7.3.8)$$

Hence using Bayes' theorem we can calculate the joint posterior distribution of $V_\varepsilon^{(k)}$ and $r_\mu^{(q)}$:

$$p(V_\varepsilon^{(k)}, r_\mu^{(q)} \mid y_t, D_{t-1}) \propto$$

$$\propto L(y_t \mid V_\varepsilon^{(k)}, r_\mu^{(q)}, D_{t-1}) \; p(V_\varepsilon^{(k)}, r_\mu^{(q)} \mid D_{t-1})$$

or using the notation of (7.3.3) we have :

$$p_t^{(kq)} = k_1 \, L(y_t \mid V_\varepsilon^{(k)}, r_\mu^{(q)}, D_{t-1}) \cdot p_{t-1}^{(kq)} \qquad (7.3.9)$$

with k_1 being a normalising factor,

$$k_1 = \sum_k \sum_q L(y_t \mid V_\varepsilon^{(k)}, r_\mu^{(q)}, D_{t-1}) \cdot p_{t-1}^{(kq)}$$

We can now derive our best estimates for the true but unknown parameters $V_{\varepsilon,t}$ and $r_{\mu,t}$ from a probability weighted combination of $V_{\varepsilon}^{(k)}$ and $r_{\mu}^{(q)}$ used by $M^{(kq)}$:

$$\hat{V}_{\varepsilon,t} \doteq E\,(V_{\varepsilon,t} \mid D_t) = \sum_k \sum_q E(V_{\varepsilon,t} \mid M^{(kq)}, D_t)\; p\,(M^{(kq)} \mid D_t)$$

$$\text{or } \hat{V}_{\varepsilon,t} = \sum_k \sum_q p_t^{(kq)} \; V_{\varepsilon}^{(k)} \tag{7.3.10}$$

and similarly

$$\hat{r}_{\mu,t} = \sum_k \sum_q p_t^{(kq)} \; r_{\mu}^{(q)} \tag{7.3.11}$$

Finally we need to produce a best estimate m_t for the true level of the process as well as an estimate C_t for its uncertainty, using information from all the models $M^{(kq)}$. That is, using the KQ normal distributions ,

$$\left\{ \begin{array}{l} (\mu_t \mid M^{(kq)}, D_t) \sim N(m_t^{(kq)}, C_t^{(kq)}) \\ \\ k = 1, 2, \ldots .\ K \quad \text{and } q = 1, 2, \ldots .\ Q \end{array} \right.$$

we want to derive estimates m_t , C_t independent of the models $M^{(kq)}$. We therefore need a single normal distribution,

$$(\mu_t \mid D_t) \sim N(m_t , C_t) .$$

which will be an approximation to the *mixed* distribution of all KQ distribution available. A similar problem was first faced in the *collapsing* procedure in chapter 3, and a solution was given in Appendix E. Using the results of Appendix E we now have :

$$m_t = \sum_k \sum_q p_t^{(kq)} \quad m_t^{(kq)} \tag{7.3.12}$$

and

$$C_t = \sum_k \sum_q p_t^{(kq)} \quad \{C_t^{(kq)} + (m_t^{(kq)} - m_t)^2\} \tag{7.3.13}$$

Hence the predictive distribution (based on all models $M^{(kq)}$) for the next time period is:

$$\left.\begin{aligned} &(y_{t+1} \mid D_t) \sim N(\hat{y}_{t+1} , \hat{Y}_{t+1}) \\ \text{where} \quad &\hat{y}_{t+1} = m_t \\ \text{and} \quad &\hat{Y}_{t+1} = C_t + \hat{V}_{\varepsilon,t} \; ({}^1/\hat{r}_{\mu,t} + 1) \end{aligned}\right\} \tag{7.3.14}$$

At this point a whole cycle of the updating process has been completed, and the procedure described in equations (7.3.2) to (7.3.14) can be repeated on receipt of every new observation.

A computer program listing for this procedure is given in Appendix H.

7.3.1. Numerical Illustration

Prior to any observation at time t = 0 our first problem is the selection of (i) values $V_\varepsilon^{(k)}$ and $r_\mu^{(q)}$ (for k = 1, 2, ...K : q = 1,=,...Q) which are fixed in each of the KQ models $M^{(kq)}$ and (ii) initial prior probabilities $\tilde{p}_o^{(kq)}$ associated with $M^{(kq)}$. If we let V_o be our initial crude estimate of V_ε then the following specification has been found to work successfully :

$$\left.\begin{array}{lll} V_\varepsilon^{(k)} = c^{k-8}\ V_o & ;\ c = 1.25;\ k = 1, 2, \ldots.K\ ; & K = 15 \\ r_\mu^{(q)} = 2^{q-1} & ;\ q = 1, 2, \ldots.Q; & Q = 8 \\ p_o^{(kq)} = {}^1/_{KQ} & ;\ k = 1, 2, \ldots.K; & q = 1, 2, \ldots.Q \end{array}\right\} \quad (7.3.1.1.)$$

The values of $(1.25)^{k-8}$ for k = 1, 2, ... 15 range from $({}^1/5)$ to 5 approximately and therefore allow for quite large errors in V_o. Similarly the $r_\mu^{(q)}$ values range from 1 to 128 and will include the true variance ratio r in almost all practical situations since this range is equivalent to expecting the optimal smoothing value α, of an EWMA to be in the range α = .08 (implied by $r_\mu^{(q)}$ = 128) to α = .62 (implied by $r_\mu^{(q)}$ = 1). Alternative specifications have also been tested but have not produced significant improvement.

As for FRM the choice of V_o is not critical to the performance of CIM and this will now be illustrated. Consider C6 Data generated with V_ε = .0025 and r_μ = 5 and let us apply CIM with V_o in error by factors of $(\frac{1}{4})$, 1 and 4, i.e.

Case 1 : $V_o = .000625$

Case 2 : $V_o = .0025$

Case 3 : $V_o = .0100$

The on line estimates $\hat{V}_{\varepsilon,t}$ and $\hat{r}_{\mu,t}$ produced by CIM for each of these three cases are tabulated in table 7.4 for times t = 4, 8, 12, ...200. It can be seen that by time t = 20 the estimates $\hat{V}_{\varepsilon,t}$ corresponding to the three cases are all reasonably close to the true V_ε value of .0025. However it takes much longer for the $\hat{r}_{\mu,t}$ estimates to converge to the true value of $r_\mu = 5$. At time t = 60, $\hat{r}_{\mu,t}$ in all cases is approximately 10 thus overestimating r_μ by a factor of 2, and by t = 140 $\hat{r}_{\mu,t}$ have converged to values very near r_μ. The slow $\hat{r}_{\mu,t}$ convergence is not however critical since as implied by the results of Appendix K, the performance of the SSM is not critically affected by estimates of r_μ which are in error by less than a factor of approximately two. That is, after t = 60 CIM can be considered to have converged near enough the optimal model for this data.

Next we illustrate the response of CIM to a discontinuity in V_ε. Consider again C6 Data this time generated with :

$$\left.\begin{aligned} V_\varepsilon &= \begin{cases} .0025 & t \leq 100 \\ .0050 & t > 100 \end{cases} \\ \text{and} \quad r_\mu &= 5 \end{aligned}\right\}$$

Table 7.4

t	case 1 $V_0 = .000625$ $\hat{V}_{\varepsilon,t}$	$\hat{r}_{\mu,t}$	case 2 $V_0 = .0025$ $\hat{V}_{\varepsilon,t}$	$\hat{r}_{\mu,t}$	case 3 $V_0 = .0100$ $\hat{V}_{\varepsilon,t}$	$\hat{r}_{\mu,t}$
4	.0004	35.51	.0013	37.11	.0047	37.65
8	.0008	32.24	.0010	36.90	.0029	44.62
12	.0014	27.47	.0016	28.73	.0028	43.29
16	.0020	42.03	.0023	43.63	.0028	49.79
20	.0022	31.13	.0027	36.25	.0031	42.48
24	.0022	22.50	.0025	26.83	.0029	33.09
28	.0021	23.75	.0023	26.04	.0026	33.28
32	.0022	30.59	.0023	32.66	.0026	39.52
36	.0021	31.47	.0022	32.76	.0025	40.27
40	.0026	26.32	.0030	34.64	.0031	36.43
44	.0027	44.49	.0032	52.58	.0032	53.04
48	.0027	36.39	.0031	44.91	.0031	45.64
52	.0026	19.08	.0030	26.36	.0031	26.99
56	.0026	8.11	.0029	10.34	.0029	10.63
60	.0026	7.89	.0028	9.38	.0029	9.64
64	.0025	7.84	.0027	8.81	.0028	9.08
68	.0025	8.34	.0027	9.20	.0027	9.44
72	.0026	8.30	.0028	9.20	.0028	9.35
76	.0025	8.00	.0026	8.55	.0027	8.75
80	.0025	8.58	.0026	9.06	.0027	9.25
84	.0026	8.71	.0027	9.27	.0027	9.43
88	.0026	9.42	.0027	9.97	.0027	10.10
92	.0026	9.74	.0027	10.25	.0027	10.38
96	.0026	10.27	.0027	10.88	.0027	10.95
100	.0026	9.78	.0026	10.16	.0026	10.27
104	.0026	9.13	.0027	9.54	.0027	9.60
108	.0025	8.94	.0026	9.19	.0026	9.28
112	.0025	7.80	.0025	7.95	.0026	8.05
116	.0025	8.12	.0025	8.23	.0025	8.35
120	.0026	8.96	.0027	9.22	.0027	9.25
124	.0028	8.96	.0029	9.71	.0029	9.71
128	.0027	8.56	.0028	9.06	.0028	9.06
132	.0028	7.53	.0029	8.06	.0029	8.07
136	.0028	6.61	.0029	6.97	.0029	6.98
140	.0027	5.69	.0028	5.87	.0028	5.88
144	.0027	5.73	.0028	5.93	.0028	5.94
148	.0027	5.77	.0027	5.94	.0027	5.96
152	.0027	5.88	.0027	6.01	.0027	6.04
156	.0027	5.97	.0027	6.10	.0027	6.13
160	.0027	6.04	.0027	6.16	.0027	6.20
164	.0027	6.20	.0028	6.34	.0028	6.37
168	.0027	6.19	.0027	6.30	.0027	6.34
172	.0026	5.88	.0027	5.95	.0027	6.00
176	.0025	4.77	.0026	4.79	.0026	4.83
180	.0025	4.02	.0025	4.03	.0025	4.04
184	.0025	3.90	.0025	3.91	.0025	3.93
188	.0024	3.92	.0025	3.91	.0025	3.94
192	.0025	3.91	.0025	3.91	.0025	3.92
196	.0025	4.03	.0025	4.03	.0026	4.03
200	.0025	4.06	.0025	4.06	.0025	4.06

$\hat{V}_{\varepsilon,t}$ and $\hat{r}_{\mu,t}$ produced by CIM when applied to C6 Data generated with $V_\varepsilon = .0025$ and $r_\mu = 5$

The on line estimates $\hat{V}_{\varepsilon,t}$; $\hat{r}_{\mu,t}$ produced by CIM and corresponding to different p_L values are tabulated in table 7.5 for t = 2,4,....200. The lower limit p_L has been included for the same reasons that it was first introduced in chapter 6 for FRM/MSM. The transformation and use of the posterior probabilities $p_{t-1}^{(kq)}$ is also analogous:

$$p_{t-1}^{(kq)*} = p_L + (1 - KQp_L)\, p_{t-1}^{(kq)} \qquad (7.3.1.2.)$$

where (7.3.1.2) is the equivalent of (6.2.5.3) given in section 6.2.5 for FRM/MSM. Given a new observation y_t the updating procedure is as described in section 7.3 except that $p_{t-1}^{(kq)*}$ is used as our prior instead of $p_{t-1}^{(kq)}$. Comparing tables 7.4 and 7.5 it can be seen that the introduction of p_L does not affect $\hat{V}_{\varepsilon,t}$ or $\hat{r}_{\mu,t}$ as long as the true parameters are constant (i.e. before t = 100). However, when a genuine change in V_ε takes place at t = 101 (where V_ε jumps to .0050 from .0025) it is clear that the on line estimates $\hat{V}_{\varepsilon,t}$ become more responsive to this change as p_L increases. Based on this and similar results from several other realisations of the C6 Data type, it has been established that $p_L = 10^{-7}$ is a good choice in the trade off between stability and speed of response. The introduction of p_L may take place whenever there are reasons to suggest that V_ε is likely to change drastically and we want FRM/SSM to respond to this change as fast as possible. If p_L is not introduced, the V_ε change will still be followed but it would take longer to converge to the new value.

Finally, CIM will be illustrated using B/J Data which with hindsight are known to be characterised by V_ε and r_μ of the order

Table 7.5

	$\rho_L = 0$		$\rho_L = 10^{-8}$		$\rho_L = 10^{-7}$		$\rho_L = 10^{-6}$	
t	$\hat{V}_{\varepsilon,t}$ $\times 10^4$	$\hat{r}_{\mu,t}$	$\hat{V}_{\varepsilon,t}$ $\times 10^4$	$\hat{r}_{\mu,t}$	$\hat{V}_{\varepsilon,t}$ $\times 10^4$	$\hat{r}_{\mu,t}$	$\hat{V}_{\varepsilon,t}$ $\times 10^4$	$\hat{r}_{\mu,t}$
2	29	33.1	29	33.1	29	33.1	29	33.1
4	14	37.2	14	37.2	14	37.2	14	37.2
6	9	41.7	9	41.7	9	41.7	9	41.7
8	11	37.5	11	37.5	11	37.5	11	37.5
10	20	33.6	20	33.6	20	33.6	20	33.6
12	16	28.9	16	28.9	16	28.9	16	28.9
14	15	37.1	15	37.1	15	37.1	15	37.1
16	23	43.6	23	43.6	23	43.6	23	43.6
18	28	49.0	28	49.0	28	49.0	28	49.0
20	27	36.3	27	36.3	27	36.3	27	36.2
22	26	41.1	26	41.1	26	41.1	26	41.1
24	25	26.8	25	26.8	25	26.8	25	26.8
26	24	24.8	24	24.8	24	24.8	24	24.8
28	23	26.0	23	26.0	23	26.0	23	26.0
30	23	31.5	23	31.5	23	31.5	24	31.5
32	23	32.7	23	32.7	23	32.7	23	32.7
34	22	33.1	22	33.1	22	33.1	22	33.1
36	22	32.8	22	32.8	22	32.8	22	32.8
38	28	43.7	28	43.7	28	43.7	28	43.7
40	30	34.6	30	34.6	30	34.6	30	34.6
42	32	43.4	32	43.4	32	43.4	32	43.4
44	32	52.6	32	52.6	32	52.6	32	52.6
46	32	51.2	32	51.2	32	51.2	32	51.2
48	31	44.9	31	44.9	31	44.9	31	44.9
50	31	30.8	31	30.8	31	30.8	31	30.8
52	30	26.4	30	26.4	30	26.4	30	26.4
54	29	12.4	29	12.4	29	12.4	30	12.4
56	29	10.3	29	10.3	29	10.3	29	10.4
58	28	10.7	28	10.7	28	10.7	28	10.8
60	28	9.4	28	9.4	28	9.4	28	9.4
62	27	8.7	27	8.7	27	8.7	27	8.7
64	27	8.8	27	8.8	27	8.8	27	8.8
66	26	8.7	26	8.7	26	8.7	26	8.8
68	27	9.2	27	9.2	27	9.2	27	9.2
70	28	9.3	28	9.3	28	9.3	29	9.3
72	28	9.2	28	9.2	28	9.2	28	9.2
74	27	9.0	27	9.0	27	9.0	27	9.0
76	26	8.5	26	8.5	26	8.5	26	8.5
78	27	8.9	27	8.9	27	8.9	27	8.9
80	26	9.1	26	9.1	26	9.1	26	9.1
82	27	9.4	27	9.4	27	9.4	27	9.4
84	27	9.3	27	9.3	27	9.3	27	9.3
86	27	9.7	27	9.7	27	9.7	27	9.7
88	27	10.0	27	10.0	27	10.0	27	10.0
90	27	10.1	27	10.1	27	10.1	27	10.1
92	27	10.3	27	10.3	27	10.3	27	10.3
94	27	10.8	27	10.8	27	10.8	27	10.9
96	27	10.9	27	10.9	27	10.9	27	10.9
98	27	10.7	27	10.7	27	10.7	27	10.8
100	26	10.2	26	10.2	26	10.2	26	10.2

CIM applied to CG Data generated with

$$V_\varepsilon = \begin{cases} .0025 & t \leq 100 \\ .0050 & t > 100 \end{cases} \quad \text{and} \quad r_\mu = 5$$

Table 7.5 cont.

102	28	10.0	28	10.0	28	10.0	28	10.0
104	29	9.6	29	9.6	29	9.7	29	9.7
106	28	8.9	28	8.9	28	8.9	28	8.9
108	28	9.2	28	9.2	28	9.2	28	9.2
110	28	8.1	28	8.1	28	8.1	28	8.1
112	28	7.2	28	7.2	28	7.2	28	7.2
114	27	7.6	27	7.6	27	7.6	27	7.6
116	27	7.7	27	7.7	27	7.7	27	7.7
118	29	7.9	29	7.9	29	7.9	29	7.9
120	32	9.4	32	9.4	32	9.4	33	9.5
122	33	9.3	33	9.3	33	9.3	34	9.4
124	38	10.2	39	10.4	46	12.1	74	19.0
126	37	9.2	37	9.3	40	10.1	59	14.9
128	37	9.1	37	9.2	38	9.6	51	12.9
130	38	7.2	41	7.4	58	8.3	83	9.7
132	38	7.7	40	7.7	49	8.3	75	10.0
134	39	7.0	40	7.1	52	7.2	77	7.6
136	38	6.3	41	6.3	56	6.1	77	5.9
138	38	5.3	41	5.2	55	4.9	73	4.4
140	37	5.0	38	5.0	47	4.7	65	4.2
142	37	5.3	39	5.2	47	5.0	64	4.5
144	37	5.0	39	5.0	47	4.7	63	4.2
146	38	5.2	39	5.2	46	4.9	61	4.4
148	37	5.0	38	5.0	43	4.8	57	4.2
150	37	5.3	38	5.2	42	5.1	55	4.5
152	37	5.2	37	5.1	40	5.0	50	4.5
154	37	5.1	38	5.1	41	5.0	52	4.4
156	38	5.3	38	5.3	41	5.2	51	4.7
158	38	5.5	39	5.5	43	5.3	54	4.8
160	38	5.4	38	5.4	40	5.3	50	4.9
162	38	5.6	38	5.6	40	5.5	49	5.2
164	39	5.7	40	5.6	44	5.5	55	5.0
166	39	5.7	40	5.6	43	5.5	54	5.1
168	39	5.6	39	5.6	41	5.5	50	5.1
170	39	5.6	39	5.6	40	5.5	48	5.2
172	38	5.2	38	5.2	40	5.1	47	4.7
174	37	4.7	37	4.7	39	4.6	46	4.1
176	37	4.1	37	4.0	39	3.9	46	3.4
178	36	3.3	37	3.3	40	2.9	46	2.5
180	36	3.4	37	3.4	41	3.1	47	2.6
182	36	3.2	36	3.2	39	2.9	45	2.4
184	36	3.3	36	3.3	39	3.0	45	2.5
186	35	3.3	36	3.3	38	3.1	43	2.6
188	35	3.3	36	3.3	38	3.1	43	2.6
190	37	3.4	37	3.3	40	3.1	46	2.6
192	37	3.3	37	3.3	40	3.0	46	2.6
194	37	3.4	38	3.4	41	3.1	46	2.7
196	38	3.5	39	3.4	43	3.1	48	2.8
198	38	3.5	39	3.4	42	3.2	47	2.8
200	38	3.5	39	3.5	42	3.2	47	2.9

of .065 and 8 respectively.

Box and Jenkins (B/J) fit an ARIMA (0,1,1) model to the complete set of 197 observations of B/J Data and their identified model implies a value for the variance ratio of $r_\mu = 7.7$. In tables 7.6 and 7.7 we tabulate the results produced by CIM using the specification given earlier by (7.3.1.1) with $V_o = .0100$. Table 7.6 shows $p_t^{(k.)}$ and $\hat{V}_{\varepsilon,t}$ for $k = 1, 2, \ldots 15$ and $t = 1, 2, \ldots 197$, where $p_t^{(k.)}$ is used to denote the marginal posterior probability corresponding to a particular $V_\varepsilon^{(k)}$ value, i.e.

$$p_t^{(k.)} = \sum_q p_t^{(kq)} \tag{7.3.1.3}$$

Similarly, table 7.7 shows $p_t^{(.q)}$ and $\hat{r}_{\mu,t}$ where $p_t^{(.q)}$ is the marginal posterior probability corresponding to $r_\mu^{(q)}$ at time t. It can be seen that by t = 197 our final on line estimate $\hat{r}_{\mu,197}$ is very close to the B/J recommended value of $r_\mu = 7.7$.

Table 7.6

$\rho_t^{(k.)}$

t	k= 1	2	3	4	5	6	7	8	9	10	11	12	13	1		
1	.09	.08	.08	.08	.08	.07	.07	.07	.07	.06	.06	.05	.05	.05		
3		.01	.02	.03	.05	.07	.09	.10	.11	.11	.10	.09	.08	.07	.	
5				.01	.03	.06	.09	.11	.13	.13	.13	.11	.09	.07	.0	
7			.01	.03	.06	.10	.14	.15	.14	.12	.09	.06	.04	.03	.02	
9				.01	.04	.07	.12	.15	.17	.15	.12	.08	.05	.03	.01	
11				.01	.04	.09	.14	.18	.18	.15	.10	.06	.03	.01		
13					.03	.07	.13	.19	.20	.16	.11	.06	.03	.01		
15				.02	.05	.10	.16	.19	.18	.14	.08	.04	.02			
17				.02	.05	.11	.18	.21	.19	.13	.07	.03	.01			
19				.02	.07	.14	.21	.22	.17	.10	.04	.02				
21				.01	.04	.11	.20	.24	.20	.11	.05	.02				.
23				.01	.05	.13	.22	.25	.19	.10	.04	.01				.0
25				.02	.07	.16	.24	.24	.15	.07	.02					.06
27				.03	.10	.20	.26	.22	.12	.05	.01					.058
29				.03	.09	.19	.26	.23	.13	.05	.01					.059
31					.05	.15	.27	.27	.17	.06	.02					.063
33					.03	.10	.23	.29	.21	.09	.03					.068
35					.03	.11	.25	.30	.20	.08	.02					.066
37					.02	.09	.24	.32	.22	.08	.02					.068
39					.02	.10	.26	.33	.20	.07	.01					.066
41					.02	.12	.28	.33	.18	.05						.064
43						.02	.10	.25	.33	.21	.07	.01				.083
45							.05	.21	.37	.26	.08	.01				.088
47							.06	.25	.38	.24	.07					.085
49							.06	.25	.38	.23	.06					.084
51							.06	.26	.40	.23	.05					.084
53							.04	.22	.40	.26	.07					.087
55							.07	.27	.39	.21	.05					.083
57							.07	.29	.40	.20	.04					.082
59							.04	.22	.42	.26	.06					.086
61							.02	.19	.43	.28	.06					.088
63							.03	.22	.44	.26	.05					.086
65								.05	.29	.44	.19	.03				.102
67								.06	.33	.44	.16	.02				.099
69								.06	.35	.43	.14	.01				.098
71								.06	.36	.43	.13	.01				.097
73								.06	.34	.44	.14	.01				.098
75								.04	.31	.47	.16	.02				.100
77								.05	.35	.46	.13	.01				.098
79								.06	.37	.45	.12					.097
81								.08	.41	.41	.09					.094
83								.10	.44	.38	.07					.092
85								.14	.48	.32	.05					.089
87							.01	.17	.50	.27	.04					.087
89							.01	.20	.52	.24	.03					.085
91							.02	.24	.52	.21	.02					.083
93							.02	.28	.51	.17	.01					.081
95							.02	.23	.53	.20	.02					.083
7							.02	.24	.52	.20	.02					.083
							.02	.27	.52	.18	.01					.082

CIM applied to B/J Data

Table 7.6 cont.

t	$P_t^{(k\cdot)}$ k = 7	8	9	10	11	$\hat{V}_{\varepsilon,t}$
101	.02	.27	.52	.17	.01	.081
103	.01	.22	.54	.21	.02	.084
105	.01	.24	.55	.19	.01	.083
107		.19	.55	.23	.02	.084
109		.21	.56	.21	.01	.084
111	.01	.24	.56	.18		.082
113	.01	.24	.57	.17		.082
115	.02	.26	.56	.16		.081
117	.02	.28	.56	.14		.081
119	.01	.25	.57	.16		.081
121	.01	.27	.57	.14		.081
123	.02	.29	.57	.13		.080
125	.02	.32	.55	.11		.079
127	.03	.37	.52	.09		.077
129	.04	.42	.48	.06		.076
131	.05	.46	.44	.05		.074
133	.04	.43	.47	.06		.075
135	.05	.46	.44	.05		.074
137	.05	.47	.43	.05		.074
139	.07	.51	.39	.03		.073
141	.08	.54	.35	.02		.071
143	.08	.55	.35	.02		.072
145	.09	.57	.31	.02		.070
147	.08	.54	.35	.02		.072
149	.10	.57	.31	.02		.070
151	.10	.58	.30	.02		.070
153	.12	.60	.26	.01		.069
155	.12	.61	.26	.01		.069
157	.13	.62	.24			.068
159	.13	.62	.24			.068
161	.14	.64	.21			.068
163	.16	.64	.18			.067
165	.18	.65	.17			.066
167	.18	.65	.16			.066
169	.20	.65	.14			.066
171	.19	.66	.15			.066
173	.17	.62	.19			.067
175	.21	.63	.15			.065
177	.15	.62	.21			.068
179	.17	.63	.19			.067
181	.17	.64	.18			.067
183	.11	.62	.26			.069
185	.12	.62	.25			.069
187	.12	.64	.23			.068
189	.14	.65	.21			.068
191	.08	.60	.31			.070
193	.09	.59	.30			.070
195	.11	.63	.26			.069
197	.09	.60	.30			.070

Table 7.7

t	$p_t^{(\cdot q)}$								$\hat{r}_{\mu,t}$
	$q=1$	2	3	4	5	6	7	8	
1	.11	.12	.12	.13	.13	.13	.13	.13	33.0
3	.17	.15	.13	.12	.11	.11	.10	.10	27.0
5	.11	.11	.12	.13	.13	.13	.14	.14	34.2
7	.09	.10	.11	.13	.14	.14	.15	.15	37.4
9	.10	.12	.13	.14	.14	.13	.13	.13	32.7
11	.07	.09	.11	.13	.14	.15	.15	.15	38.2
13	.06	.09	.12	.15	.16	.15	.14	.13	35.4
15	.10	.14	.17	.18	.15	.11	.08	.07	22.6
17	.07	.11	.16	.19	.17	.13	.10	.07	25.3
19	.06	.11	.17	.21	.18	.13	.08	.05	21.3
21	.04	.08	.14	.18	.19	.16	.12	.09	29.3
23	.04	.08	.14	.20	.21	.16	.11	.07	26.4
25	.05	.10	.17	.22	.21	.14	.07	.04	20.0
27	.05	.12	.20	.25	.21	.11	.05	.02	15.3
29	.05	.10	.17	.23	.21	.14	.07	.03	18.5
31	.03	.07	.16	.25	.25	.16	.06	.02	18.6
33	.03	.09	.19	.28	.25	.13	.04		14.9
35	.03	.07	.16	.27	.27	.15	.05	.01	16.4
37	.02	.06	.15	.25	.28	.17	.06	.01	18.5
39	.02	.05	.14	.25	.29	.18	.06	.01	18.4
41	.02	.05	.14	.27	.30	.17	.05		17.1
43	.02	.06	.12	.21	.26	.20	.09	.03	22.8
45		.01	.06	.17	.29	.27	.14	.04	29.6
47		.02	.07	.19	.30	.26	.13	.04	28.4
49		.02	.07	.20	.31	.25	.12	.04	27.0
51		.01	.05	.17	.30	.28	.14	.04	29.9
53		.01	.07	.21	.33	.25	.10	.03	25.7
55		.02	.11	.29	.33	.18	.05	.01	18.5
57		.02	.10	.28	.36	.19	.05		18.7
59		.01	.07	.22	.34	.25	.09	.02	23.9
61			.05	.21	.36	.26	.09	.02	24.0
63			.06	.23	.38	.24	.06	.01	21.2
65			.03	.14	.32	.32	.15	.04	31.4
67			.03	.14	.32	.32	.15	.04	31.7
69			.02	.13	.32	.33	.15	.04	32.2
71			.02	.12	.31	.34	.16	.05	33.1
73			.03	.13	.32	.32	.15	.04	31.8
75			.01	.10	.29	.35	.19	.06	36.3
77			.02	.10	.30	.35	.18	.05	35.3
79			.02	.12	.32	.34	.16	.04	32.6
81			.02	.15	.35	.32	.13	.03	28.8
83			.04	.20	.40	.27	.08	.01	23.3
85			.06	.28	.42	.20	.04		18.4
87			.06	.30	.44	.18	.03		17.1
89			.06	.30	.45	.17	.02		16.7
91			.07	.34	.44	.13	.01		15.0
93			.07	.36	.45	.11			14.3
95			.06	.32	.46	.15	.01		15.6
97			.10	.36	.40	.12	.01		14.3
99			.11	.39	.38	.10			13.6

CIM applied to B/J Data

Table 7.7 cont.

t	$p_t^{(\cdot q)}$						$\hat{r}_{\mu,t}$
	q = ...	3	4	5	6	7	
101		.10	.37	.40	.11	.01	14.1
103		.07	.35	.43	.13	.01	15.1
105		.07	.35	.43	.13	.01	14.9
107		.07	.33	.44	.15	.02	15.7
109		.06	.32	.45	.15	.02	16.0
111		.06	.33	.45	.15	.01	15.8
113		.05	.31	.46	.16	.02	16.2
115		.07	.35	.43	.13	.01	15.2
117		.06	.33	.45	.15	.01	15.8
119		.05	.31	.45	.16	.02	16.2
121		.05	.30	.46	.17	.02	16.6
123		.05	.30	.46	.17	.02	16.5
125		.06	.33	.45	.14	.01	15.6
127		.08	.39	.42	.10		13.9
129		.09	.43	.40	.08		12.8
131		.10	.46	.37	.05		11.9
133		.08	.42	.41	.08		13.1
135		.09	.42	.41	.08		13.0
137		.12	.46	.36	.06		12.1
139		.12	.48	.34	.05		11.6
141		.13	.50	.32	.04		11.1
143		.12	.50	.33	.05		11.3
145		.12	.50	.33	.04		11.2
147		.13	.49	.32	.05		11.3
149		.14	.51	.30	.04		10.9
151		.13	.51	.31	.04		11.0
153		.16	.54	.27	.03		10.2
155		.14	.53	.29	.03		10.6
157		.14	.53	.30	.03		10.7
159		.13	.53	.30	.03		10.7
161		.13	.53	.30	.03		10.7
163		.13	.53	.31	.03		10.7
165		.12	.54	.31	.03		10.7
167		.12	.53	.31	.03		10.8
169		.14	.54	.29	.03		10.4
171		.12	.53	.32	.03		10.9
173	.02	.21	.53	.22	.02		9.4
175	.02	.25	.55	.17	.01		8.6
177	.01	.19	.52	.24	.03		9.7
179	.01	.20	.53	.24	.02		9.6
181	.01	.19	.53	.24	.02		9.7
183		.15	.52	.29	.03		10.5
185	.01	.16	.51	.28	.03		10.4
187		.15	.52	.29	.03		10.5
189		.15	.52	.29	.03		10.5
191		.12	.50	.32	.05		11.2
193	.01	.19	.53	.24	.02		9.7
195	.01	.20	.55	.21	.02		9.2
197	.01	.17	.52	.26	.03		10.0

7.4. On line estimation of V_ε and r_μ when both are assumed to be constant.

The methods proposed so far, for on line estimation of V_ε or joint estimation of V_ε and r_μ allow these true process parameters to change through time and respond reasonably well even when the noise variance V_ε has a discontinuity at a particular point in time. When V_ε and r_μ are assumed to be constant however, it is possible to design a joint estimation method based on the limiting properties of the SSM which is computationally much more efficient than CIM. Because of the constant variance assumption we will refer to this new method as CVM which stands for "Constant Variance Method".

Leonard and Harrison (L/H) [30] recently proposed a method for the joint estimation of V_ε and r_μ in a SSM, assuming continuous distributions for the unknown variances V_ε and V_μ. This however leads to a rather cumbersome and computationally lengthy updating system requiring a number of one dimensional numerical integrations to be performed at each point in time. The choice for the number of intervals used in the numerical integrations is 1000 although this may be reduced by first detecting the region in which the posterior distribution of the variance ratio r_μ is highly concentrated. Cantarelis and Johnston [10] have however demonstrated that CVM, assuming an 8-point discrete probability distribution for r_μ, produces practically identical forecasting performance while being computationally much more efficient. Comparisons of the two methods will be made in section 7.4.5 using B/J Data.

7.4.1. Optimality conditions for the likelihood function

Consider a steady state process with constant true parameters V_ε and r_μ . Then at time t the likelihood of a SSM operating with fixed initially nominated estimates $V_{\varepsilon,N}$ and $r_{\mu,N}$ being the correct model for the process, is given by the following log-likelihood function L*.

$$L^* = \sum_t \log L_t \qquad (7.4.1.1)$$

$$\text{where } L_t \propto p(y_t \mid V_{\varepsilon,N}, r_{\mu,N}, D_{t-1}) \qquad (7.4.1.2)$$

When the model is in its limiting form (see section 2.3 of chapter 2) then an expression for the expected value of $\log L_t$ can be derived as shown in Appendix K :

$$E^* = E(\log L_t) \propto -\tfrac{1}{2}\left\{\log\left[\frac{V_{\varepsilon,N}}{1-A_N}\right] + \left[\frac{1-A_N}{V_{\varepsilon,N}}\right]\left[\frac{2A_N V_\varepsilon + V_\mu}{A_N(2-A_N)}\right]\right\}$$

Clearly maximisation of L* is in the limit equivalent to maximisation of E* and therefore differentiating the latter with respect to $V_{\varepsilon,N}$ and A_N and equating to zero, results in *optimality conditions* I and II.

$$\frac{\partial E^*}{\partial V_{\varepsilon,N}} = 0 \Longrightarrow V_{\varepsilon,N} = (1-A_N)\left[\frac{2A_N V_\varepsilon + V_\mu}{A_N(2-A_N)}\right] \qquad (\mathrm{I})$$

$$\frac{\partial E^*}{\partial A_N} = 0 \Longrightarrow A_N = \frac{-1 + [1 + 4r_\mu]^{\frac{1}{2}}}{2r_\mu} \qquad (\mathrm{II})$$

Note that the notation here is consistent with that always used for a SSM and which was first introduced in section 2.3.

The expression for $\dot{V}_{\varepsilon,N}$ given by (I) above, is the maximum likelihood value of V_ε given A_N. That is given $r_{\mu,N}$ which automatically fixes A_N (see section 2.3) then the optimal value of V_ε is $V_{\varepsilon,N}$ as given by (I). This result is essential for CVM as will be seen in the next section which describes how CVM works.

7.4.2. CVM basic concepts

Let M* denote a SSM operating with a fixed nominated variance ratio $r_{\mu,N}$ and an on line estimate of V_ε, $\hat{V}_{\varepsilon,t}$, in place of $V_{\varepsilon,N}$

where

$$\hat{V}_{\varepsilon,t} = (1-A_N)\ (\mathrm{MSE})_t \qquad (7.4.2.1)$$

with $$(\mathrm{MSE})_t = (^1/_t) \sum_{i=1}^{t} e_i^2 \qquad (7.4.2.2)$$

Note that (i) A_N is known initially being a function of $r_{\mu,N}$ (see section 2.3) which is fixed in M* and (ii) e_i is the one step ahead forecast error at time $t = i$.

As time progresses $(\mathrm{MSE})_t$ tends to the expected value of the squared error which has been derived in Appendix D:

$$\lim_{t \to \infty} (\mathrm{MSE})_t = E(e_t^2) = \frac{2A_N V_\varepsilon + V_\mu}{A_N (2-A_N)}$$

Hence using this result and (7.4.2.1) we can write :

$$\underset{t \to \infty}{\text{limit}} \hat{V}_{\varepsilon,t} = (1 - A_N) \left[\frac{2A_N V_\varepsilon + V_\mu}{A_N (2-A_N)} \right] \qquad (7.4.2.3)$$

But the right hand side (7.4.2.3) is the maximum likelihood estimate of V_ε given $r_{\mu,N}$, that is $V_{\varepsilon,N}$ as given by optimality condition I.

If $r_{\mu,N}$ happened to be equal to the true variance ratio r_μ then A_N would also satisfy optimality condition II and the limiting form of the model M* would be optimal. In practice however r_μ is unknown and hence we need a procedure for its on line estimation. This is an ideal situation for the application of a class I approach similar to the one used in CIM earlier.

It is assumed that the uncertainty about the correct model and its parameters can be adequately described by a set of Q alternative models $M^{(q)}$ for q = 1, 2, ...Q. Model $M^{(q)}$ operates in exactly the same way as model M* described above with $r_\mu^{(q)}$ in place of $r_{\mu,N}$. Each model is assigned an associated initial prior probability $p_0^{(q)}$ which is updated with every new observation using Bayesian relationships, so that at time t, $p_t^{(q)}$ reflects the posterior probability of $M^{(q)}$ being the correct model for the process under observation. Assuming that the range of $r_\mu^{(q)}$ is chosen sufficiently large to ensure that r_μ lies within it, then this class I formulation will converge to the correct variance ratio r_μ thus satisfying optimality condition II. Since optimality condition I is automatically satisfied within each model as previously described in terms of M^*, it follows that a probability weighted linear combination of $M^{(q)}$ will also converge to the correct model.

7.4.3. CVM updating procedure

Let the prior information at time t required by the updating procedure be of the following form for each model $M^{(q)}$; $q = 1, 2, \ldots Q$:

$$(\mu_{t-1} \mid M^{(q)}, D_{t-1}) \sim N(m_{t-1}^{(q)}, C_{t-1}^{(q)}) \tag{7.4.3.1}$$

$$p(M^{(q)} \mid D_{t-1}) = p_{t-1}^{(q)} \tag{7.4.3.2}$$

$$E(V_\varepsilon \mid M^{(q)}, D_{t-1}) = \hat{V}_{\varepsilon,t-1}^{(q)} \tag{7.4.3.3}$$

As soon as y_t becomes available it is used together with $m_{t-1}^{(q)}$, $C_{t-1}^{(q)}$, $p_{t-1}^{(q)}$, $\hat{V}_{\varepsilon,t-1}^{(q)}$ to produce posterior estimates $m_t^{(q)}$, $C_t^{(q)}$, $p_t^{(q)}$ $\hat{V}_{\varepsilon,t}^{(q)}$ so that the procedure can be repeated on receipt of y_{t+1} etc.

The updating procedure used is now described.

(i) Updating $\hat{V}_{\varepsilon,t-1}^{(q)}$

Prior to y_t the best estimate of V_ε available in model $M^{(q)}$ is $\hat{V}_{\varepsilon,t-1}^{(q)}$ and if this together with (7.4.3.1) are used in the Kalman filter equations (see (2.2.5) in chapter 2) for a SSM, we get the following for each $M^{(q)}$; $q = 1, 2, \ldots Q$:

$$\left.\begin{aligned}
\hat{y}_t^{(q)} &= m_{t-1}^{(q)} \\
R_t^{(q)} &= C_{t-1}^{(q)} + \hat{V}_{\varepsilon,t-1}^{(q)} / r_\mu^{(q)} \\
\hat{Y}_t^{(q)} &= R_t^{(q)} + \hat{V}_{\varepsilon,t-1}^{(q)} \\
A_t^{(q)} &= R_t^{(q)} / \hat{Y}_t^{(q)}
\end{aligned}\right\} \qquad (7.4.3.4.)$$

As soon as y_t becomes available the maximum likelihood estimate of V_ε conditional on D_t and $M^{(q)}$ is given by equation (7.4.2.1) which in the limit was shown to satisfy optimality condition I. Hence,

$$V_{\varepsilon,t}^{(q)} = (1 - A_q) \quad (MSE)_t^{(q)} \qquad (7.4.3.5)$$

where $$(MSE)_t^{(q)} = (^1/_t) \sum_{i=1}^{t} (e_i^{(q)})^2 \qquad (7.4.3.6)$$

$$e_t^{(q)} = y_t - \hat{y}_t^{(q)}$$

and $$A_q = \left\{ - 1 + [\ 1 + 4r_{\mu,q}\ ]^{\frac{1}{2}} \right\} \Big/ 2r_{\mu,q} \qquad (7.4.3.7)$$

(ii) <u>Updating of $p_{t-1}^{(q)}$</u> :

Bayes' theorem can be used to write:

$$p(M^{(q)} \mid D_t) = p(y_t \mid M^{(q)}, D_{t-1}) \; p(M^{(q)} \mid D_{t-1})$$

or using (7.4.3.2) and (7.4.3.4) ,

$$p_t^{(q)} \propto L_t^{(q)} \quad p_{t-1}^{(q)}$$

where $L_t^{(q)} \propto [\hat{Y}_t^{(q)}]^{-\frac{1}{2}} \quad \exp \{-\tfrac{1}{2} [e_t^{(q)}]^2 / \hat{Y}_t^{(q)}\}$

(iii) Updating of $m_{t-1}^{(q)}$ and $C_{t-1}^{(q)}$

Given (7.4.3.1), (7.4.3.5) and using the Kalman filter equations (7.4.3.4) it follows that the posterior distribution of μ_t conditional on D_t and $M^{(q)}$ is:

$$(\mu_t \mid M^{(q)}, D_t) \sim N(m_t^{(q)}, C_t^{(q)})$$

where $m_t^{(q)} = m_{t-1}^{(q)} + A_t^{(q)} \quad e_t^{(q)}$

and $C_t^{(q)} = A_t^{(q)} \quad \hat{V}_{\varepsilon,t}^{(q)}$

This completes the updating cycle since all the prior information assumed by (7.4.3.1),(7.4.3.2) and (7.4.3.3) have been updated using the latest observation y_t. A computer program listing for this updating procedure is given in Appendix H.

7.4.4. Unconditional posterior estimates and prediction

It is now possible to combine the information available from all Q models in order to derive unconditional (not dependent on $M^{(q)}$) posterior estimates for V_ε, r_μ, μ_t and the variance of μ_t. These are derived using similar arguments as for CIM and can be written as follows:

$$\hat{V}_{\varepsilon,t} = E(V_\varepsilon \mid D_t) = \sum_q p_t^{(q)} \hat{V}_{\varepsilon,t}^{(q)}$$

$$\hat{r}_{\mu,t} = E(r_\mu \mid D_t) = \sum_q p_t^{(q)} r_\mu^{(q)}$$

$$m_t = E(\mu_t \mid D_t) = \sum_q p_t^{(q)} m_t^{(q)}$$

$$C_t = \mathrm{Var}\,(\mu_t \mid D_t) = \sum_q p_t^{(q)} \{C_t^{(q)} + (m_t^{(q)} - m_t)^2\}$$

Finally the above estimates can be used for prediction. The one step ahead predictive distribution is for example,

$$(y_{t+1} \mid D_t) \sim N\,(y_{t+1}, \hat{Y}_{t+1})$$

where $\hat{y}_{t+1} = m_t$

and $\hat{Y}_{t+1} = C_t + V_{\varepsilon,t}\,({}^1/\hat{r}_{\mu,t} + 1)$

7.4.5. Numerical Illustration

Initially at time $t = 0$ CVM requires a set of values to be specified :

(i) Q : the number of $M^{(q)}$ models used.

(ii) $r_{\mu}^{(q)}$: the variance ratio fixed in model $M^{(q)}$.

and (iii) $p_{o}^{(q)}$: initial prior probability associated with $M^{(q)}$

The following default choice for these values has been found satisfactory :

$$Q = 8$$

$$r_{\mu}^{(q)} = 2^{q-1} \; ; \quad q = 1, 2, \ldots Q \qquad (7.4.5.1)$$

$$p_{o}^{(q)} = {}^{1}/_{Q}$$

Alternative specifications have also been tried but have lead to insignificant improvement of forecasting performance and are therefore not reported. One such case where $Q = 100$, that is using 100 models $M^{(q)}$ instead of eight, is reported in [10].

Two extra practical suggestions are:

a) At time $t = 1$ the forecast errors $e_{t}^{(q)}$ used in $(MSE)_{t}^{(q)}$ could be artificially set equal to an arbitrarily small value in order to ensure that unusually large errors $e_{1}^{(q)}$ resulting from bad initial estimates of the process level μ_o, are not included in $(MSE)_t^{(q)}$ which would then affect the uptdating of $\hat{V}_{\varepsilon,t}$.

b) At time t = 0 complete ignorance is assumed by each model $M^{(q)}$ about V_ε and therefore it would be reasonable to prevent the updating of our prior views about the variance ratio until a better V_ε estimate is available. Usually ten points are sufficient for $\hat{V}_{\varepsilon,t}$ to converge to a reasonable estimate of V_ε and therefore the following procedure is suggested :

$$p_t^{(q)} = p_o^{(q)} \qquad \text{for } t = 1, 3, \ldots .10$$

but for t > 10 , $p_t^{(q)}$ is updated in the usual way described in 7.4.3. It should be stressed here that CVM converges to maximum likelihood estimates of V_ε and r_μ , and the above practical refinements are only designed to accelerate the speed of convergence.

CVM will now be illustrated using (i) C6 Data and (ii) B/J Data.

In table 7.8 we have tabulated the on line estimates $\hat{V}_{\varepsilon,t}$ and $\hat{r}_{\mu,t}$ produced by CVM when applied on C6 Data generated with V_ε = .0025 and r = 25. The default choice given by (7.4.5.1) has been used for the starting values required by the method. It can be seen that the $\hat{V}_{\varepsilon,t}$ estimates converge to the true value V_ε much faster than $\hat{r}_{\mu,t}$ converges to r_μ . The reason for this is that the likelihood function (see Appendix K) is

Table 7.8

t	$\hat{V}_{\varepsilon,t}$	$\hat{r}_{\mu,t}$	t	$\hat{V}_{\varepsilon,t}$	$\hat{r}_{\mu,t}$
2	.0002	31.9	102	.0025	53.6
4	.0002	28.0	104	.0025	48.1
6	.0002	40.2	106	.0025	44.4
8	.0006	15.3	108	.0025	47.4
10	.0012	55.0	110	.0024	43.6
12	.0010	55.6	112	.0024	39.2
14	.0012	49.9	114	.0024	41.5
16	.0018	74.1	116	.0024	43.3
18	.0021	75.6	118	.0025	42.7
20	.0021	67.6	120	.0026	47.3
22	.0020	74.6	122	.0026	44.1
24	.0021	69.4	124	.0027	50.6
26	.0019	66.8	126	.0027	50.4
28	.0018	64.4	128	.0027	50.8
30	.0018	69.7	130	.0028	50.9
32	.0018	69.2	132	.0028	52.8
34	.0017	70.3	134	.0028	51.1
36	.0017	70.1	136	.0027	49.4
38	.0021	80.6	138	.0027	42.8
40	.0025	77.2	140	.0027	36.0
42	.0025	81.0	142	.0027	36.5
44	.0025	86.6	144	.0027	38.3
46	.0024	87.6	146	.0026	36.1
48	.0024	87.4	148	.0026	33.7
50	.0025	83.0	150	.0026	35.5
52	.0025	85.0	152	.0026	35.1
54	.0026	70.9	154	.0026	34.3
56	.0026	66.4	156	.0026	33.9
58	.0026	71.8	158	.0026	34.7
60	.0026	61.5	160	.0026	35.0
62	.0025	55.7	162	.0026	35.4
64	.0025	52.6	164	.0026	35.2
66	.0025	50.0	166	.0026	34.6
68	.0025	53.1	168	.0026	34.3
70	.0027	52.9	170	.0026	34.8
72	.0026	53.1	172	.0025	34.1
74	.0025	53.1	174	.0025	32.7
76	.0025	50.5	176	.0025	29.2
78	.0025	51.0	178	.0025	22.7
80	.0024	48.5	180	.0025	22.3
82	.0025	55.3	182	.0025	20.8
84	.0025	55.1	184	.0025	20.6
86	.0025	56.7	186	.0024	19.7
88	.0025	57.5	188	.0024	20.0
90	.0025	57.2	190	.0025	21.3
92	.0025	57.8	192	.0025	21.1
94	.0025	60.3	194	.0025	21.8
96	.0025	59.1	196	.0025	22.3
98	.0025	59.2	198	.0025	22.0
100	.0024	56.5	200	.0025	22.5

CVM applied to C6 Data generated

with $V_\varepsilon = .0025$ $r_\mu = 25$

very insensitive to the variance ratio and consequently it is to be expected that in order to identify it accurately, a large number of observations is required.

Finally consider B/J Data and let us apply CVM again using the default choice for the starting values. The results are tabulated in table L.1 of Appendix L where for each time period from t = 1 to t = 197 the following information are given:

$$y_t \ , \ \hat{y}_t \ , \ e_t, \ m_t, \ C_t \ , \ \hat{Y}_t, \ \hat{V}_{\varepsilon,t} \ , \ \hat{r}_{\mu,t}$$

where e_t is the system forecast error at time t and all other variables are as defined in section 7.4.4. Note that initially, complete ignorance has been assumed for both V_ε and r_μ. The final posterior estimate for V_ε is .071 which is close to those implied by B/J (.071) and L/H (.066). The final posterior estimate of 9.25 for r_μ is much nearer that reported by B/J (7.7) than the L/H estimate of 5.0. These differences however are not critical because of the insensitivity of the likelihood function to the variance ratio.

7.5. Concluding remarks

Four methods of on line variance estimation in the SSM have been proposed.

The first two, FRM/SSM and VRM/SSM, are simply the methods first introduced and fully described in chapter 6 for the MSM, modified for use in a SSM. They both assume an initially fixed estimate ($r_{\mu,N}$) of the true variance ratio r_μ and produce on line estimates of V_ε given $r_{\mu,N}$. It was shown that their accuracy is expected to increase as r_μ becomes larger and that they are generally appropriate for EWMA processes where the optimal smoothing constant α is less than approximately .27. Their properties are very similar to those described in chapter 6 but slightly different recommendations for p_L ($p_L = 10^{-5}$ instead of 10^{-4} is recommended) and β ($\beta = 1.3$ instead of 1.4 is recommended) are made. This is a consequence of the fact that the SSM is not as robust as MSM to forecast errors of the order of 2σ or larger.

CIM was then described for on line joint estimation of V_ε and r_μ, and is particularly applicable to EWMA processes with r_μ smaller than approximately 10 (I.e. $\alpha > .27$). It was shown that the initial estimate of the noise variance, V_o, whilst it results in different $V_\varepsilon^{(k)}$ values its effect on the on line estimates $\hat{V}_{\varepsilon,t}$ and $r_{\mu,t}$ is insignificant. The introduction of a lower limit p_L (with a recommended value of $p_L = 10^{-7}$) was shown to accelerate the speed of response to a possible abrupt change in V_ε without affecting the effectiveness of on line estimation during periods of relative stability in V_ε and r_μ.

Another procedure, CVM, was proposed as a joint V_ε and r_μ on line estimation method, assuming that these true parameters are absolutely constant through time. This method is computationally much more efficient than CIM, which is mainly the result of two facts:

(i) The first optimality condition (I) is satisfied within each model $M^{(q)}$ and therefore only a one dimensional class I approach is required to satisfy the second optimality condition (II). Thus CVM uses Q models $M^{(q)}$ in contrast to the KQ models $M^{(kq)}$ required by CIM.

(ii) only a very small number of models (Q = 8) is sufficient for satisfactory on line identification of the correct model.

Finally, it must be stressed that all these methods described for the SSM can be modified for a LGM which is characterised by $V_{\varepsilon,N}$ and w_N, the latter parameter being an initially fixed nominated estimate of the EWR discount factor w. This estimate (w_N) then determines the other variances $V_{\mu,N}$ and $V_{\beta,N}$ in LGM. In the same way that $r_{\mu,N}$ is of secondary importance in the SSM, w_N also would not be expected to be critical to the performance of the LGM relative to V_ε, and therefore FRM and VRM may be used to estimate the noise variance. Similarly CIM and CVM can be used for joint estimation of V_ε and w in an analogous way that they were used to estimate V_ε and r_μ in the SSM.

CHAPTER 8

ALTERNATIVE FORMULATIONS OF THE MSM

8.1. Foreword

Several alternative formulations of the four state MSM (described fully in chapter 3) have been examined and three of them will be proposed in this chapter :

1) The 1PS (1 - point system)
2) The 2PS (2 - point system)
3) The 33S ("three-three" system)

Our objectives are twofold :

(i) to improve computational efficiency,

and (ii) to improve the performance of the MSM and particularly its speed of response to abrupt growth changes without at the same time spoiling the stability of the system during quiet periods.

In chapter 5 it was seen that a faster growth response could be achieved by spoiling the balance of the Π's in the standard SSP. This improvement was however achieved at the expense of significant instability over a period of "no change" and therefore our aim in this chapter is to operate with the standard SSP but using different formulations of the MSM.

8.2. Expansion of matrix operations

The derivation and presentation of the equations and recurrence relationships for the MSM as described in chapter 3, is clear and concise in matrix notation and of course completely general. From a computational point of view it is inefficient to perform general matrix operations (which is very convenient using a computer program language such as BASIC) as all the elements in the $\underline{F}_t$ and $\underline{G}$ matrices are either zero or one. For example, consider the one step ahead forecast equation for the LGM of section 2.4:

$$\hat{y}_t = \underline{F}_t \, \underline{G} \, \underline{m}_{t-1} \qquad (8.2.1)$$

Substituting for $\underline{F}_t$, $\underline{G}$, $\underline{m}_{t-1}$ their LGM forms,

$$\underline{F}_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \quad \underline{G} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \underline{m}_{t-1} = \begin{bmatrix} m_{t-1} \\ b_{t-1} \end{bmatrix}$$

it can be seen that (8.2.1) is equivalent to the following algebraic equation :

$$\hat{y}_t = m_{t-1} + b_{t-1} \qquad (8.2.2)$$

thus reducing the 6 multiplications and 3 additions to one operation. These sort of savings can be made on all the *Kalman filter* and *collapsing process* equations leading to a decrease in CPU time by a factor of the order of 30.

The value of opening up the matrices becomes even greater with more complex models. For instance, consider the linear growth seasonal model described in C.1 (Appendix C). The one step ahead forecast equation is still that given above by (8.2.1) but the $\underline{F}_t$ $\underline{G}$ and $\underline{m}_{t-1}$ matrices and vectors now contain components relevant to seasonality:

$$\hat{y}_t = \underline{F}_t \, \underline{G} \, \underline{m}_{t-1} =$$

$$= \underset{1 \times 14}{\left[\begin{array}{cc:cc} 1 & 0 & 1 & 0 \ldots 0 \end{array} \right]} \; \underset{14 \times 14}{\left[\begin{array}{cc:cccccc} 1 & 1 & & & & & & \\ 0 & 1 & & & \underline{0} & & & \\ \hdashline & & 0 & 1 & 0 & & & \\ & & & 0 & 1 & 0 & & \\ & & & & & \ddots & & \\ \underline{0} & & & & & & & \\ & & & & & 0 & 0 & 1 \\ & & 1 & 0 & \ldots & \ldots & \ldots & 0 \end{array} \right]} \; \underset{14 \times 1}{\left[\begin{array}{c} m_{t-1} \\ b_{t-1} \\ \hdashline S_{t+11} \\ S_t \\ S_{t+1} \\ \vdots \\ S_{t+9} \\ S_{t+10} \end{array} \right]}$$

or $\hat{y}_t = m_{t-1} + b_{t-1} + S_t$ (8.2.3) where m_{t-1} = best estimate of the level at time t - 1

b_{t-1} = " " " " growth " " "

S_t = " " " " current seasonal effect.

Clearly if $\hat{y}_t$ is calculated using matrix operations then 210 multiplications and 195 additions will be required instead of only 2 additions which are strictly necessary if the matrix operations have been expanded to produce the algebraic equation given by (8.2.3).

The cost savings are clearly significant and although these procedures have been employed in our computer programs (see Appendix H) all descriptions of the operations have been left in vector and matrix form for clarity and generality.

8.3. The 1 - Point System (1PS)

As pointed out in chapter 2 and illustrated in Appendix C the *principle of superposition* can be used to build Bayesian systems of any degree of generality. Suppose for example that the process parameter vector $\underline{\theta}_t$ consists of 13 seasonal factors, 13 promotional factors (see section C.2 of Appendix C) as well as level and growth. Then the MSM handles vectors and matrices of order 28. Under such circumstances or when the MSM is to be used for forecasting a large number of items, the computational and storage requirements may well be unacceptably high for many potential users of the system. In view of this possible constraint, 1PS has been designed to minimise the computational effort without affecting the performance of the MSM.

The improvement in computational efficiency is achieved by excluding a number of unlikely state transitions $i \to j$ assuming that no more than one discontinuity (i.e. occurrence of states 2, 3, 4) can occur in two consecutive time periods. That is, if the process is in an "outlier", "growth change" or "step change" state at time t-1 then it will be in a "no change" state at time t. This assumption implies a transition probability matrix shown in figure 8.1 together with that implicitly used by the MSM:

FIGURE 8.1

MSM

i \ j	1	2	3	4
1	$\pi^{(1)}$	$\pi^{(2)}$	$\pi^{(3)}$	$\pi^{(4)}$
2	"	"	"	"
3	"	"	"	"
4	"	"	"	"

1PS

i \ j	1	2	3	4
1	$\pi^{(1)}$	$\pi^{(2)}$	$\pi^{(3)}$	$\pi^{(4)}$
2	1	-	-	-
3	1	-	-	-
4	1	-	-	-

Hence only 7 $i \rightarrow j$ transitions are considered by 1PS instead of 16 which leads to a considerable reduction in computational effort, especially when the process parameter vector $\underline{\Theta}_t$ incorporates a large number of parameters in addition to the usual process level μ_t and growth β_t. The two major areas where computation is reduced are a) in the *Kalman updating* procedure where 7 units of CPU are now required instead of 16 and b) in the *collapsing process* where the CPU is reduced by a factor of four as a result of the fact that instead of 16, only 4 state transitions (correspinding to the first row of the 1PS transition matrix) require *collapsing* at each point in time.

Apart from this difference in the number of state transitions considered, the 1PS works like the MSM and as will be shown in the next section its performance is practically identical to that of the MSM except for step changes which are better modelled by the 1PS.

8.3.1. Comparison of 1PS and MSM

Consider first the *performance measures* $R^{(j)}$ (j = 1, 2, 3, 4), as defined in chapter 4 and used extensively in chapter 5. The performance produced by the MSM and 1PS assuming a known noise variance (i.e. $V_{\varepsilon,N} = V_\varepsilon = .0025$) and using the standard SSP is illustrated by tabulating their MSE and z responses in table 8.1 and graphing their system growth and level responses ($R^{(3)}$, $R^{(4)}$) in figures 8.1a and 8.1b respectively.

Table 8.1

	$R^{(1)}$ MSE	$R^{(2)}$ z response to outliers of size : 4σ	10σ	20σ
MSM	100	.3	.3	.7
1PS	100	.3	.3	.7

It can be seen that the 1PS behaves in a practically identical way to the MSM except for step changes of medium or large size (i.e. 5σ or larger) which are recognised slightly earlier by 1PS. This is because the 1PS formulation assumes that no more than one discontinuity can occur over two consecutive time periods. A better understanding of these findings can be achieved by a careful examination of the transition probabilities $p^{(ij)}$ (first defined in section 3.3) produced by the two systems over a set of data containing a quiet period as well as all the different types of discontinuities implied by states 2, 3, and 4. An artificially generated series has therefore been constructed by introducing an outlier of size 10σ,

$R^{(3)}$ RESPONSE
½ σ GROWTH CHANGE
1PS
MSM
1σ GROWTH CHANGE
1PS
MSM
1·5σ GROWTH CHANGE
1PS
MSM

FIGURE 8.1a

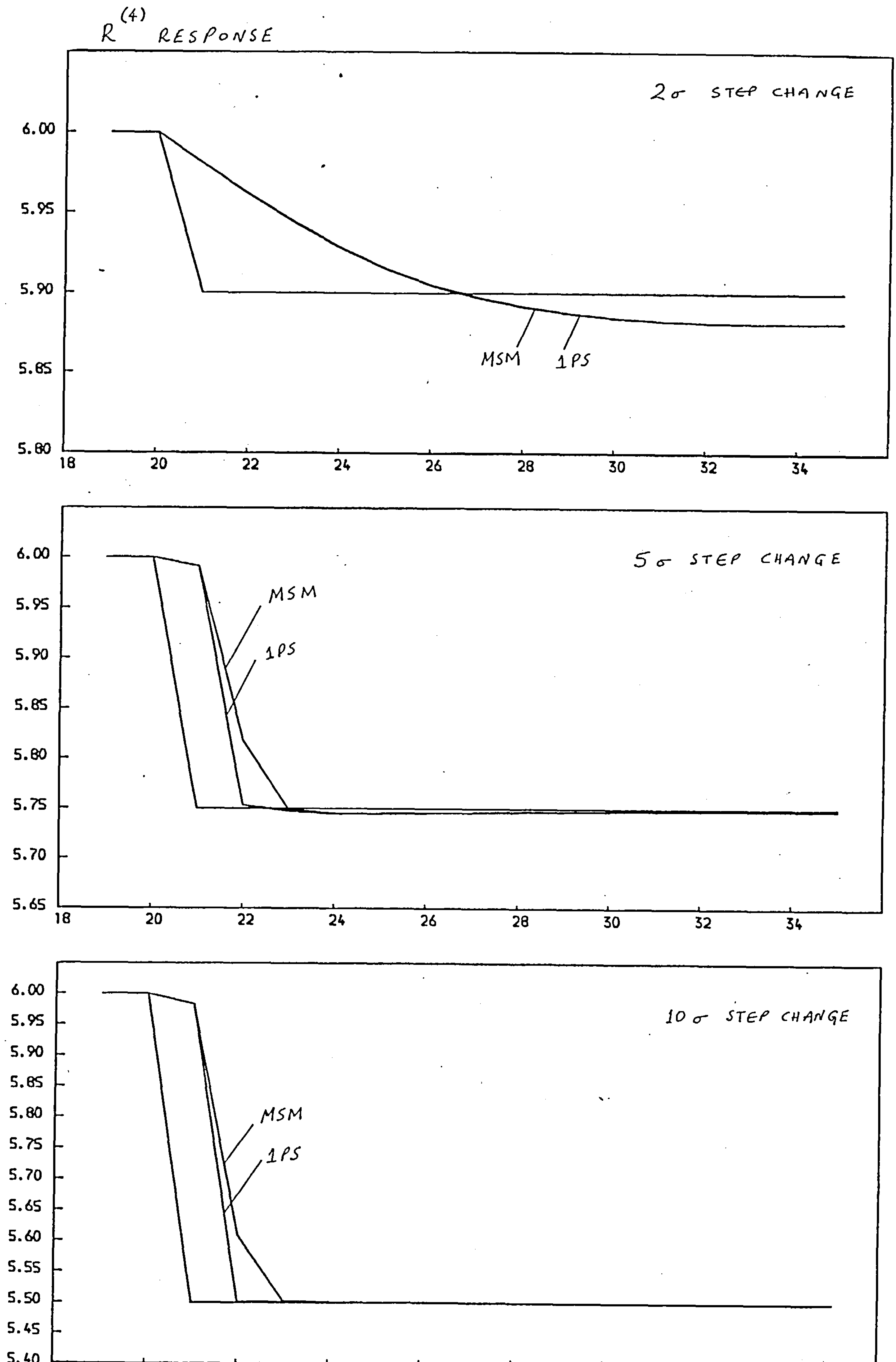

$R^{(4)}$ RESPONSE
2σ STEP CHANGE
MSM
1PS
5σ STEP CHANGE
MSM
1PS
10σ STEP CHANGE
MSM
1PS
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
5.60
5.55
5.50
5.45
5.40
18
20
22
24
26
28
30
32
34

FIGURE 8.1b

a step change of size 5σ and a growth change of size 1σ, on the C6 Data (as defined in chapter 6, section 6.2.1) at times t = 21, t = 41, and t = 61 respectively. The resulting series will from now on be referred to as C8 Data (standing for Chapter 8 Data) and is graphed in figure 8.2 together with the one step ahead forecasts produced by 1PS. The equivalent forecasts produced by the MSM are also shown although they are almost identical at all times except for a few periods just after the step change at t = 41 when 1PS responds faster. The transition probabilities $p^{(ij)}$ for times t = 1, 2, ...100 produced by MSM and 1PS (using $V_{\varepsilon,N} = V_{\varepsilon}$ and the standard SSP) when applied to C8 Data are tabulated in tables 8.2 and 8.3 respectively. Recall that (i) $p^{(ij)}$ at time t stands for the probability of a transition from STATE i at time t - 1 to STATE j at time t given y_t and (ii) i and j may take a value of 1, 2, 3 or 4 such that :

state 1 implies a "no change" state
" 2 implies an "outlier" "
" 3 implies a "growth change" state
" 4 implies a "step change" "

At time t = 21 for example we know (see figure 8.2) that an "outlier" has occurred and hence the transition probability $p^{(12)}$ at time t = 21 is expected to be high. This is confirmed in table 8.2 where $p^{(12)}$ at t = 21 is .943 and in table 8.3 where the corresponding value is .969.

FIGURE 8.2

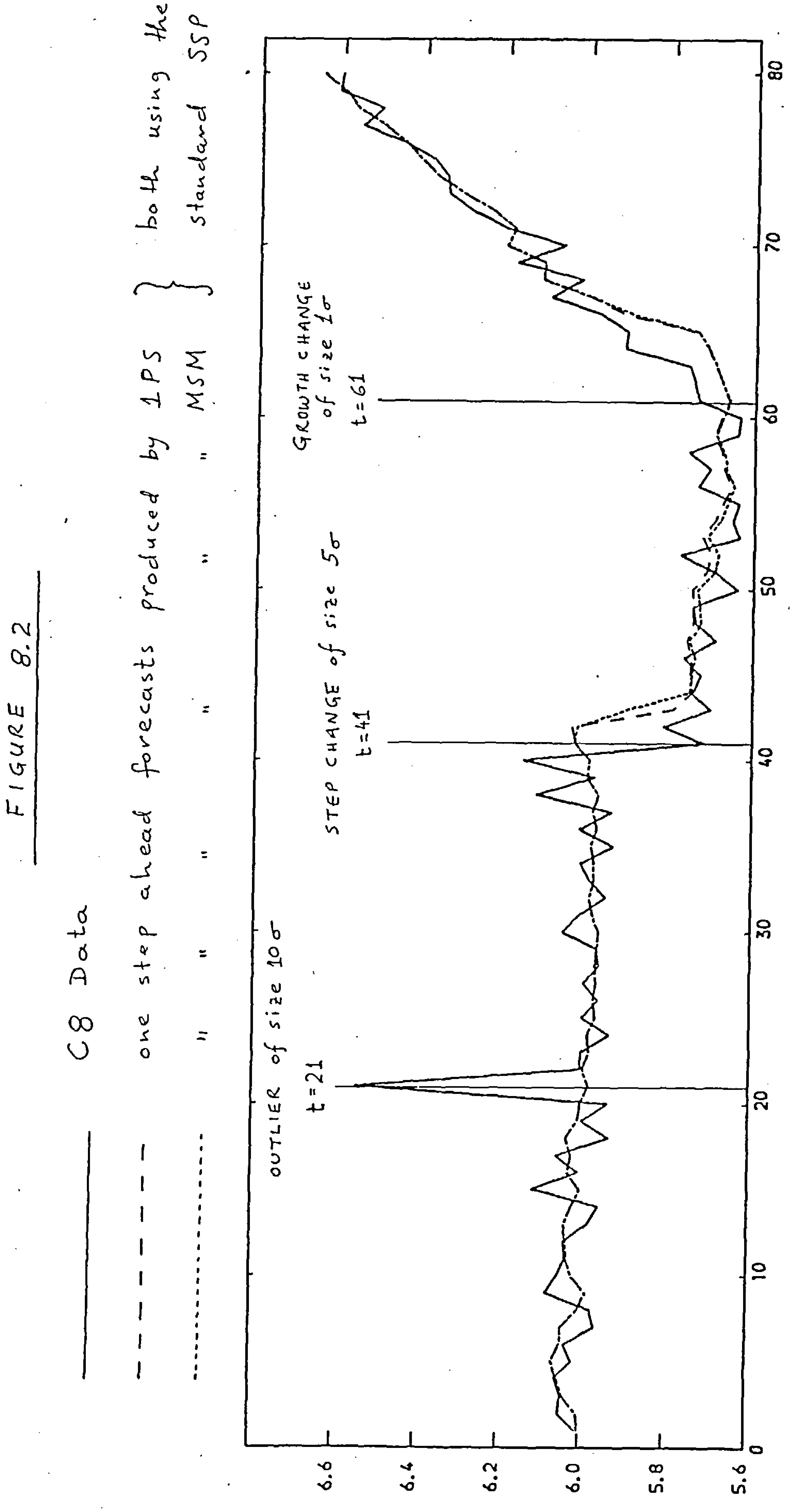


FIGURE 8.2
C8 Data
one step ahead forecasts produced by 1PS } both using the
" " " " " MSM } standard SSP
OUTLIER of size 10σ
t=21
STEP CHANGE of size 5σ
t=41
GROWTH CHANGE of size 1σ
t=61
6.6
6.4
6.2
6.0
5.8
5.6
0
10
20
30
40
50
60
70
80

$p^{(ij)} \times 10^3$ produced by the MSM when applied to C8 Data

t	$p^{(11)}$	$p^{(21)}$	$p^{(31)}$	$p^{(41)}$	$p^{(12)}$	$p^{(22)}$	$p^{(32)}$	$p^{(42)}$	$p^{(13)}$	$p^{(23)}$	$p^{(33)}$	$p^{(43)}$	$p^{(14)}$	$p^{(24)}$	$p^{(34)}$	$p^{(44)}$
1	876	91	3	3	21	2	0	0	3	0	0	0	1	0	0	0
2	960	15	3	1	18	0	0	0	3	0	0	0	1	0	0	0
3	965	13	2	1	15	0	0	0	3	0	0	0	0	0	0	0
4	967	12	2	0	15	0	0	0	3	0	0	0	0	0	0	0
5	963	13	2	0	18	0	0	0	3	0	0	0	1	0	0	0
6	967	13	2	1	14	0	0	0	3	0	0	0	0	0	0	0
7	952	13	3	0	27	0	0	0	3	0	0	0	1	0	0	0
8	962	16	3	1	14	0	0	0	3	0	0	0	0	0	0	0
9	933	20	3	0	38	1	0	0	4	0	0	0	1	0	0	0
10	955	23	3	1	14	1	0	0	3	0	0	0	0	0	0	0
11	970	13	2	0	12	0	0	0	2	0	0	0	0	0	0	0
12	971	11	2	0	12	0	0	0	2	0	0	0	0	0	0	0
13	963	12	2	0	18	0	0	0	3	0	0	0	1	0	0	0
14	957	13	3	1	21	0	0	0	3	0	0	0	1	0	0	0
15	875	39	2	0	74	2	0	0	5	0	0	0	2	0	0	0
16	901	80	2	1	12	1	0	0	2	0	0	0	0	0	0	0
17	967	13	2	0	14	0	0	0	3	0	0	0	0	0	0	0
18	919	20	3	0	51	1	0	0	4	0	0	0	2	0	0	0
19	940	41	2	1	11	1	0	0	2	0	0	0	0	0	0	0
20	959	11	2	0	22	0	0	0	3	0	0	0	1	0	0	0
21	0	0	0	0	943	23	3	1	0	0	0	0	30	1	0	0
22	0	984	0	0	0	13	0	0	0	3	0	0	0	0	0	0
23	974	8	2	0	12	0	0	0	3	0	0	0	0	0	0	0
24	963	13	2	0	17	0	0	0	3	0	0	0	1	0	0	0
25	963	19	2	0	13	0	0	0	3	0	0	0	0	0	0	0
26	970	13	2	0	12	0	0	0	2	0	0	0	0	0	0	0
27	970	12	2	0	13	0	0	0	3	0	0	0	0	0	0	0
28	971	13	2	0	12	0	0	0	2	0	0	0	0	0	0	0
29	972	11	2	0	11	0	0	0	2	0	0	0	0	0	0	0
30	947	12	3	0	33	0	0	0	4	0	0	0	1	0	0	0
31	954	26	3	1	13	0	0	0	3	0	0	0	0	0	0	0
32	967	14	2	0	14	0	0	0	3	0	0	0	0	0	0	0
33	970	14	2	0	11	0	0	0	2	0	0	0	0	0	0	0
34	971	11	2	0	13	0	0	0	3	0	0	0	0	0	0	0
35	962	14	2	0	18	0	0	0	3	0	0	0	1	0	0	0
36	960	20	2	0	15	0	0	0	3	0	0	0	0	0	0	0
37	963	16	2	0	15	0	0	0	3	0	0	0	0	0	0	0
38	688	20	5	0	264	4	1	0	8	0	0	0	8	0	0	0
39	701	281	2	1	9	3	0	0	2	1	0	0	0	0	0	0
40	645	19	7	1	306	4	1	0	8	0	0	0	10	0	0	0
41	0	1	0	0	639	314	6	8	0	0	0	0	20	10	0	0
42	14	254	2	358	0	344	0	12	0	3	0	1	0	11	0	0
43	900	4	8	14	45	22	0	1	3	0	0	0	1	1	0	0
44	959	19	2	2	13	1	0	0	3	0	0	0	0	0	0	0
45	970	11	2	0	13	0	0	0	3	0	0	0	0	0	0	0
46	968	13	2	0	14	0	0	0	3	0	0	0	0	0	0	0
47	958	15	2	0	20	0	0	0	3	0	0	0	1	0	0	0
48	963	18	2	1	13	0	0	0	3	0	0	0	0	0	0	0
49	970	12	2	0	13	0	0	0	3	0	0	0	0	0	0	0
50	936	17	3	0	38	1	0	0	4	0	0	0	1	0	0	0

Table 8.2

51	952	29	2	1	12	0	0	0	2	0	0	0	0	0	0	0
52	942	14	3	0	35	0	0	0	4	0	0	0	1	0	0	0
53	908	58	2	0	26	1	0	0	3	0	0	0	1	0	0	0
54	961	19	3	1	13	0	0	0	3	0	0	0	0	0	0	0
55	969	12	2	0	13	0	0	0	3	0	0	0	0	0	0	0
56	942	17	2	0	33	0	0	0	4	0	0	0	1	0	0	0
57	955	22	3	1	14	0	0	0	3	0	0	0	0	0	0	0
58	944	11	4	1	34	1	0	0	4	0	0	0	1	0	0	0
59	932	44	2	0	17	1	0	0	3	0	0	0	1	0	0	0
60	964	14	3	1	15	0	0	0	3	0	0	0	0	0	0	0
61	947	19	2	0	26	0	0	0	3	0	0	0	1	0	0	0
62	951	17	4	1	22	1	0	0	3	0	0	0	1	0	0	0
63	955	15	4	1	21	0	0	0	3	0	0	0	1	0	0	0
64	156	2	30	3	759	16	3	1	5	0	0	0	24	1	0	0
65	592	45	37	205	23	88	1	3	2	1	0	1	1	3	0	0
66	958	7	4	4	20	2	0	0	3	0	0	0	1	0	0	0
67	944	13	4	1	34	1	0	0	3	0	0	0	1	0	0	0
68	899	60	2	1	32	1	0	0	3	0	0	0	1	0	0	0
69	930	42	2	1	21	1	0	0	3	0	0	0	1	0	0	0
70	829	63	3	0	94	2	0	0	5	0	0	0	3	0	0	0
71	903	75	2	2	14	1	0	0	3	0	0	0	0	0	0	0
72	963	12	2	1	18	0	0	0	3	0	0	0	1	0	0	0
73	966	13	2	1	15	0	0	0	3	0	0	0	0	0	0	0
74	967	14	2	0	13	0	0	0	3	0	0	0	0	0	0	0
75	969	12	2	0	14	0	0	0	3	0	0	0	0	0	0	0
76	970	13	2	0	12	0	0	0	2	0	0	0	0	0	0	0
77	964	12	2	0	18	0	0	0	3	0	0	0	1	0	0	0
78	951	23	2	0	20	0	0	0	3	0	0	0	1	0	0	0
79	963	20	2	0	12	0	0	0	2	0	0	0	0	0	0	0
80	967	12	2	0	15	0	0	0	3	0	0	0	0	0	0	0
81	883	23	3	0	80	1	0	0	5	0	0	0	3	0	0	0
82	925	52	4	2	13	1	0	0	3	0	0	0	0	0	0	0
83	969	13	2	0	13	0	0	0	3	0	0	0	0	0	0	0
84	971	12	2	0	12	0	0	0	2	0	0	0	0	0	0	0
85	956	11	3	0	25	0	0	0	3	0	0	0	1	0	0	0
86	950	29	2	0	15	0	0	0	3	0	0	0	0	0	0	0
87	957	17	2	0	19	0	0	0	3	0	0	0	1	0	0	0
88	959	21	2	0	14	0	0	0	3	0	0	0	0	0	0	0
89	937	18	2	0	36	1	0	0	4	0	0	0	1	0	0	0
90	947	36	2	1	12	0	0	0	2	0	0	0	0	0	0	0
91	972	12	2	0	12	0	0	0	2	0	0	0	0	0	0	0
92	966	12	2	0	16	0	0	0	3	0	0	0	1	0	0	0
93	957	19	2	0	18	0	0	0	3	0	0	0	1	0	0	0
94	922	26	2	0	42	1	0	0	4	0	0	0	1	0	0	0
95	927	52	2	0	15	1	0	0	3	0	0	0	0	0	0	0
96	965	13	2	1	15	0	0	0	3	0	0	0	0	0	0	0
97	969	15	2	0	11	0	0	0	2	0	0	0	0	0	0	0
98	971	11	2	0	13	0	0	0	3	0	0	0	0	0	0	0
99	968	14	2	0	13	0	0	0	3	0	0	0	0	0	0	0
100	970	12	2	0	13	0	0	0	3	0	0	0	0	0	0	0

Table 8.2 cont.

$p^{(ij)} \times 10^3$ produced by the 1PS when applied to C8 Data

t	$p^{(11)}$	$p^{(21)}$	$p^{(31)}$	$p^{(41)}$	$p^{(12)}$	$p^{(13)}$	$p^{(14)}$
1	869	101	3	3	21	3	1
2	961	15	3	1	18	3	1
3	964	14	3	1	15	3	0
4	966	13	2	1	15	3	0
5	962	14	2	0	18	3	1
6	966	14	2	1	14	3	0
7	951	14	3	1	27	3	1
8	961	18	3	1	14	3	0
9	932	21	3	0	38	4	1
10	954	24	3	1	14	3	0
11	969	14	2	0	12	2	0
12	970	13	2	0	12	2	0
13	962	13	2	0	18	3	1
14	956	14	4	1	21	3	1
15	874	41	3	0	74	5	2
16	896	86	2	1	12	2	0
17	966	14	2	0	14	3	0
18	918	22	3	0	51	4	2
19	937	45	2	1	11	2	0
20	959	12	3	0	22	3	1
21	0	0	0	0	969	0	31
22	0	1000	0	0	0	0	0
23	984	0	0	0	13	3	0
24	962	15	2	0	17	3	1
25	962	20	2	0	13	3	0
26	969	14	2	0	12	2	0
27	969	13	2	0	13	3	0
28	970	14	2	0	11	2	0
29	972	12	2	0	11	2	0
30	946	12	3	0	33	4	1
31	951	29	3	1	13	3	0
32	966	15	2	0	14	3	0
33	969	15	2	0	11	2	0
34	970	12	2	0	13	3	0
35	961	16	2	0	17	3	1
36	959	21	2	0	15	3	0
37	962	18	2	0	15	3	0
38	685	21	5	0	272	8	9
39	680	306	2	1	9	2	0
40	642	12	5	0	322	8	10
41	0	2	0	0	967	0	31
42	50	70	1	877	1	0	0
43	965	0	0	0	30	3	1
44	959	22	2	1	13	3	0
45	969	13	2	0	13	3	0
46	969	14	2	0	12	3	0
47	956	15	2	0	23	3	1
48	961	22	2	1	12	2	0
49	971	12	2	0	12	2	0
50	917	14	4	1	57	5	2

Table 8.3

51	938	43	3	1	12	2	0
52	953	15	2	0	25	3	1
53	914	45	2	0	34	4	1
54	954	23	3	1	15	3	0
55	966	13	2	1	15	3	0
56	946	20	2	0	27	3	1
57	958	22	3	1	13	3	0
58	949	12	4	1	30	4	1
59	933	44	2	0	18	3	1
60	962	15	3	1	16	3	1
61	947	22	2	0	25	3	1
62	951	18	5	1	21	3	1
63	955	16	4	1	20	3	1
64	161	2	33	4	771	5	25
65	648	47	42	235	24	2	1
66	962	10	3	1	20	3	1
67	944	14	4	1	33	3	1
68	895	65	2	1	32	3	1
69	926	46	2	1	21	3	1
70	826	69	3	0	94	5	3
71	899	80	2	2	14	2	0
72	964	11	3	1	18	3	1
73	965	14	3	1	15	3	0
74	966	16	2	0	13	3	0
75	968	13	2	0	14	3	0
76	969	14	2	0	12	2	0
77	963	13	2	0	18	3	1
78	949	25	2	0	20	3	1
79	962	21	2	0	12	2	0
80	966	13	2	0	15	3	0
81	883	25	3	0	80	5	3
82	922	56	4	2	13	3	0
83	969	13	2	0	13	3	0
84	970	14	2	0	12	2	0
85	955	12	3	0	25	3	1
86	948	32	2	0	15	3	0
87	956	19	2	0	19	3	1
88	957	23	2	0	14	3	0
89	937	20	2	0	36	4	1
90	944	39	2	1	12	2	0
91	971	13	2	0	12	2	0
92	965	14	2	0	16	3	1
93	956	20	2	0	18	3	1
94	921	29	2	0	42	4	1
95	924	56	2	0	15	3	0
96	965	13	3	1	15	3	0
97	968	16	2	0	11	2	0
98	970	12	2	0	13	3	0
99	967	15	2	0	13	3	0
100	969	13	2	0	13	3	0

Table 8.3 cont.

At time $t = 41$ when a step change occurs both MSM and 1PS treat the actual observation as an outlier but with an important difference. The MSM has $p^{(12)} = .639$ and $p^{(22)} = .314$, $(p^{(12)} + p^{(22)} = .953)$ in contrast to 1PS which has $p^{(12)} = .967$ with $p^{(22)}$ implicitly zero since unlikely transitions such as $2 \rightarrow 2$ are excluded. Given an additional observation at $t = 42$ the 1PS can easily infer that the previous observation was not an outlier but a step change and this is reflected in a high $p^{(41)}$ which from table 8.3 can be seen to be .877. The MSM however can not make the same inference with such confidence due to the extra uncertainty associated with the $2 \rightarrow 2$ transition. Hence at time $t = 42$ the MSM has a $p^{(41)}$ as low as .358 (compared with .877 of 1PS) which explains why its response to step changes is slower than that of the 1PS.

From table 8.2 it can also be seen that the sum of transition probabilities $p^{(22)}$, $p^{(32)}$, $p^{(42)}$, $p^{(23)}$, $p^{(33)}$, $p^{(43)}$, $p^{(24)}$, $p^{(34)}$ and $p^{(44)}$ is very small (for all t) compared to the sum of the remaining probabilities. It follows that the contribution of state transitions (22), (32), (42), (23), (33), (43), (24), (34) and (44) to the performance of the MSM is relatively insignificant and this explains why the 1PS which excludes these unlikely transitions produces a very similar response at all times except at step changes which are recognised earlier by the latter system.

Finally the sensitivity of 1PS has been examined to errors in the nominated variance $V_{\varepsilon,N}$. The 1PS has been found to behave in an almost identical manner to the MSM when the latter operates with a $V_{\varepsilon,N}$

in some error from V_ε (with the exception of the slightly faster response to step changes). The results of the last section in chapter 5 testing the robustness of the MSM are therefore also valid for the 1PS. It can be considered as a substitute for the MSM and on the computer used the CPU is reduced by a factor of approximately 2. However, the 1PS has not achieved the second of the stated objectives namely to increase the speed of response to growth changes.

8.4. The 2 - Point System (2PS)

Consider the state of a process after the observation at time t - 1. It was seen in section 3.3 that the MSM characterises the process evolution in terms of four models $M_{t-1}^{(i)}$ with associated probabilities $p_{t-1}^{(i)}$. Before y_t becomes known each model $M_{t-1}^{(i)}$ summarises the prior information about the parameter vector in terms of a normal distribution:

$$(\underline{\Theta}_{t-1} \mid D_{t-1}, M_{t-1}^{(i)}) \sim N(\underline{m}_{t-1}^{(i)}, \underline{C}_{t-1}^{(i)}) \qquad (8.4.1)$$

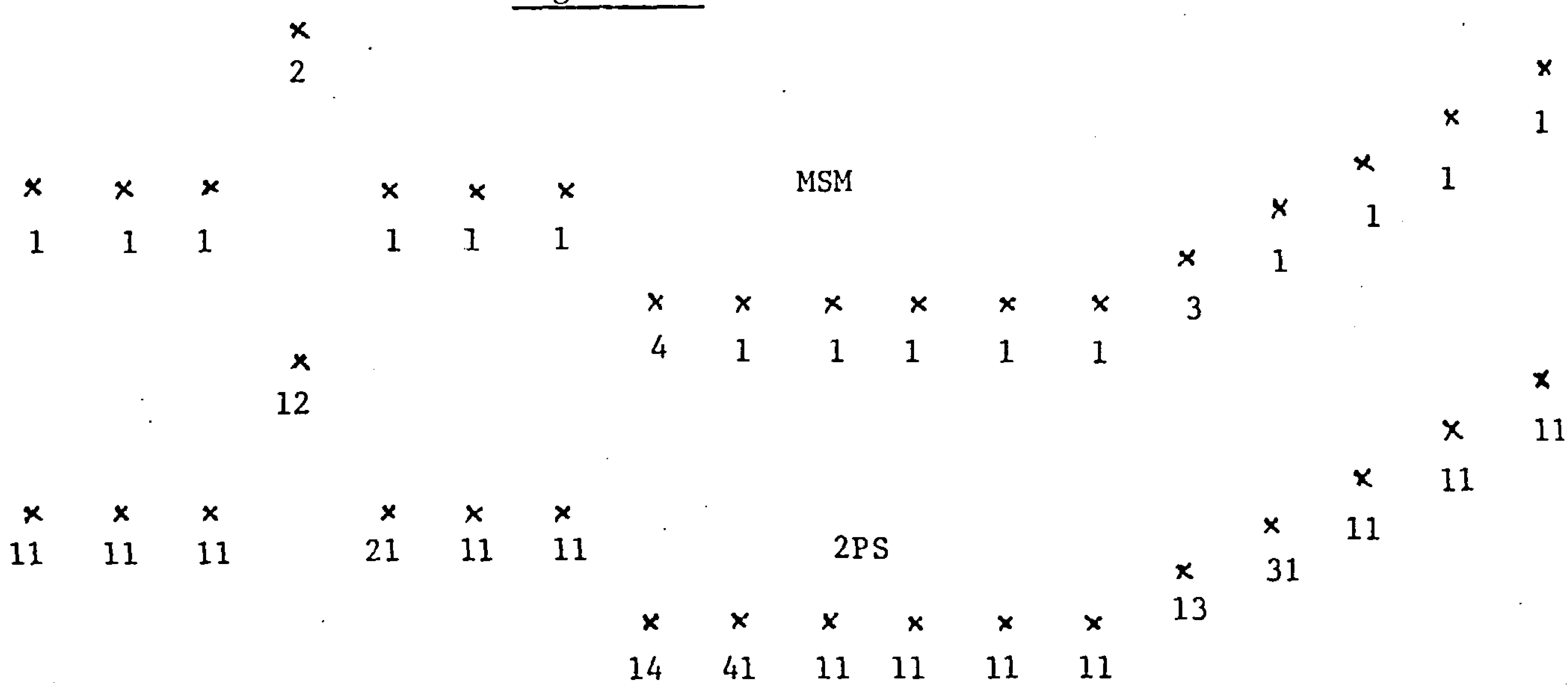
As soon as the observation at time t (y_t) becomes available, the likelihoods of state transitions $i \rightarrow j$ (i, j = 1, 2, 3, 4) are assessed and the information associated with these $i \rightarrow j$ transitions are *collapsed* to produce (i) $p_t^{(j)}$ reflecting the probability of the process currently being in state j and (ii) a new prior distribution for $\underline{\Theta}_t$ in the same form as (8.3.1) :

$$(\underline{\Theta}_t \mid D_{t-1}, M_t^{(j)}) \sim (N(\underline{m}_t^{(j)}, \underline{C}_t^{(j)}) \qquad (8.4.2)$$

Hence the MSM views the process at any point in time t, as being in one of four states 1, 2, 3 or 4 without any reference of the process state at time t - 1.

The fundamental difference of 2PS from MSM is that it defines a state over two consectutive points in time t-1, t. The 2PS can for example differentiate between the process "being in a no change state at time t having come from a no change state at time t-1" and "being in a no change at time t having come from a growth change state at time t-1". Hence the *states* of 2PS at time t are of the form ij (instead of just j as in the MSM) and there are clearly 16 possible states for i, j = 1, 2, 3, 4. In the following figure an attempt is made to illustrate the different ways used by MSM and 2PS to view the underlying process. A purely deterministic process is used for clarity.

<u>Figure 8.3</u>



Of course certain ij states can be excluded as being unrealistic, and the 2PS (in line with 1PS) assumes that if at any point in time t the process is in a "change" state then at t+1 the process is back in a "no change" state. It follows that the 7 i → j *state transitions* implied by the 1PS probability transition matrix of figure 8.1, will now become the 7 *states* of 2PS. If we further assume that over three successive points in time at most one "change" state can occur, then the following state transition probability matrix can be used for the 2PS:

FIGURE 8.4

$$
\begin{array}{cc}
ij \nearrow jk & \begin{array}{ccccccc} 11 & 12 & 13 & 14 & 21 & 31 & 41 \end{array} \\
\begin{array}{c} 11 \\ 12 \\ 13 \\ 14 \\ 21 \\ 31 \\ 41 \end{array} &
\left[\begin{array}{ccccccc}
\Pi^{(1)} & \Pi^{(2)} & \Pi^{(3)} & \Pi^{(4)} & * & * & * \\
* & * & * & * & 1 & * & * \\
* & * & * & * & * & 1 & * \\
* & * & * & * & * & * & 1 \\
1 & 0 & 0 & 0 & * & * & * \\
1 & 0 & 0 & 0 & * & * & * \\
1 & 0 & 0 & 0 & * & * & *
\end{array} \right]
\end{array}
$$

2PS probability transition matrix

A * implies that such a transition is not possible by definition.

A 0 implies that such a transition is possible but has been excluded as unrealistic.

$\Pi^{(j)}$ $j = 1, 2, 3, 4$ are the same as for MSM and the standard SSP values can be used.

Hence in moving from time t to t+1 there are 10 possible transitions of the form ij → jk as implied by the probability matrix of figure 8.4. A state transition ij → jk will from now on be denoted by ijk for simplicity. Given this fundamental difference in the definition of a *state*, the 2PS works in exactly the same way as described in chapter 3 for the MSM. We will therefore only briefly outline the *Kalman updating* and *Collapsing* procedures used by 2PS pointing out where it differs from the MSM equivalent, given in section 3.3.

8.4.1. 2PS Kalman updating and Collapsing

At time t prior to observing y_{t+1}, the 2PS characterises the process evolution in terms of seven models denoted by $M_t^{(ij)}$ in contrast to the four $M_t^{(j)}$ models in the MSM. Each model $M_t^{(ij)}$ summarises the prior information about the parameter vector and process state in terms of (i) and (ii) below :

(i) $$(\underline{\Theta}_t \mid D_t, M_t^{(ij)}) \sim N(\underline{m}_t^{(ij)}, \underline{C}_t^{(ij)}) \qquad (8.4.1.1)$$

where $m_t^{(ij)}$ and $\underline{C}_t^{(ij)}$ have an obvious and analogous meaning to $m_t^{(j)}$ and $\underline{C}_t^{(j)}$ of the MSM.

(ii) $$\left.\begin{array}{l} p_t^{(ij)} = \text{probability that the process is currently} \\ \qquad \text{in state (ij).} \end{array}\right\} \qquad (8.4.1.2)$$

Given that the process is in state (ij) at time t the 2PS considers 10 possible transition models $M^{(ijk)}$ for a process transition ij → jk. The state transition probabilities assumed by 2PS have been given in figure 8.4.

As soon as y_{t+1} becomes available, ten Normal distributions

$$(\underline{\Theta}_{t+1} \mid D_t , M^{(ijk)}) \sim N(\underline{m}^{(ijk)} , \underline{C}^{(ijk)}) \qquad (8.4.1.3)$$

associated with transitions ij → jk can be derived by applying the set of Kalman filter equations given by (3.3.7) with the following simple notation differences:

1) The subscript t-1 now becomes t
2) The superscript (i) now becomes (ij)
3) The superscript (ij) now becomes (ijk)
4) The superscript (j) now becomes (jk)

We can then easily arrive at $p^{(ijk)}$ reflecting the probability of $M^{(ijk)}$ being the correct model, in the same way that $p^{(ij)}$ was derived in equation (3.3.9).

Finally the information associated with $M^{(ijk)}$, that is $p^{(ijk)}$ and $\underline{m}^{(ijk)}$, $\underline{C}^{(ijk)}$ are collapsed as shown in figure 8.5. (analogous to figure 3.3) and the set of equations below which are analogous to (3.3.12).

$$\left.\begin{aligned}
p_{t+1}^{(jk)} &= \begin{cases} p^{(1jk)} & \text{for } jk = 12,\ 13,\ 14,\ 21,\ 31,\ 41 \\ p^{(1jk)} + p^{(2jk)} + p^{(3jk)} + p^{(4jk)} & \text{for } jk = 11 \end{cases} \\
k_i &= p^{(i11)} \,/\, p_{t+1}^{(11)} \qquad \text{for } i = 1,\ 2,\ 3,\ 4. \\
m_{t+1}^{(jk)} &= \begin{cases} m^{(1jk)} & \text{for } jk = 12,\ 13,\ 14,\ 21,\ 31,\ 41 \\ \sum\limits_i k_i \ m^{(ijk)} & \text{for } i = 1,\ 2,\ 3,\ 4 \text{ and } jk = 11 \end{cases} \\
b_{t+1}^{(jk)} &= \begin{cases} b^{(1jk)} & \text{for } jk = 12,\ 13,\ 14,\ 21,\ 31,\ 41 \\ \sum\limits_i k_i \ b^{(ijk)} & \text{for } i = 1,\ 2,\ 3,\ 4 \text{ and } jk = 11 \end{cases} \\
c_{11,\ t+1}^{(jk)} &= \begin{cases} c_{11}^{(1jk)} & \text{for } jk = 12,\ 13,\ 14,\ 21,\ 31,\ 41 \\ \sum k_i \ \{c_{11}^{(ijk)} + [m^{(ijk)} - m_{t+1}^{(jk)}]^2\} & \\ \qquad \text{for } i = 1,\ 2,\ 3,\ 4 \text{ and } jk = 11 & \end{cases} \\
c_{12,\ t+1}^{(jk)} &= \begin{cases} c_{12}^{(1jk)} & \text{for } jk = 12,\ 13,\ 14,\ 21,\ 31,\ 41 \\ \sum\limits_i k_i \left\{ c_{12}^{(ijk)} + [m^{(ijk)} - m_{t+1}^{(jk)}]\,[b^{(ijk)} - b_{t+1}^{(jk)}] \right\} & \\ \qquad \text{for } i = 1,2,\ 3,\ 4 \text{ and } jk = 11 & \end{cases} \\
c_{22,\ t+1}^{(jk)} &= \begin{cases} c_{22}^{(1jk)} & \text{for } jk = 12,\ 13,\ 14,\ 21,\ 31,\ 41 \\ \sum\limits_i k_i \ \{c_{22}^{(ijk)} + [b^{(ijk)} - b_{t+1}^{(jk)}]^2\} & \\ \qquad \text{for } i = 1,\ 2,\ 3,\ 4 \text{ and } jk = 11 & \end{cases}
\end{aligned}\right\} \quad (8.4.1.4)$$

Collapsing in the 2PS

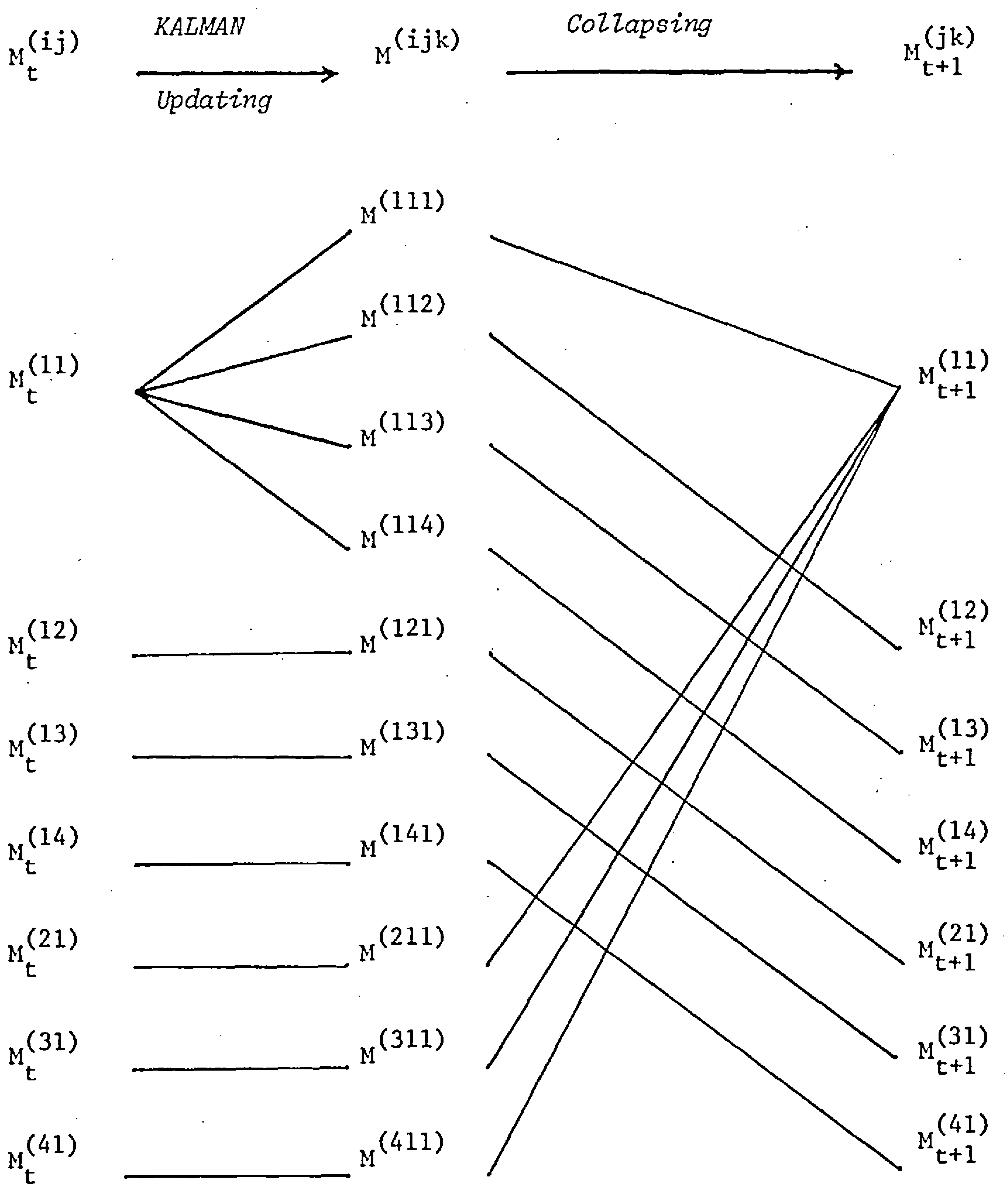


Figure 8.5.

8.4.2. Numerical Illustration of 2PS

Let us first look at the response of 2PS on C8 Data (as defined in 8.3.1 and graphed in figure 8.2) in terms of:

(i) transition probabilities $p^{(ijk)}$ at time t

(ii) 2PS system estimate of the level, m_t, at time t.

and (iii) 2PS system estimate of the growth, b_t, at time t.

These are shown in table 8.4 together with the MSM estimates of the level and growth, for comparisons. Recall that in table 8.2 in section 8.2.1, the MSM probabilities were given and can therefore be compared with those of table 8.4.

The response of 2PS at the points of discontinuities is particularly interesting. AT t = 21 (when an outlier occurs) the probability of a transition to state 12 is high ($p^{(112)}$ = .969 from table 8.4) which implies that the freak observation is virtually ignored as can be seen from m_t and b_t at t = 21. Given the next observation at t = 22, the system can infer with certainty ($p^{(121)}$ = 1.000 correct to 3 decimal places) that the process has moved from state 12 to state 21 and similarly at t = 23 $p^{(211)}$ = 1.000.

At t = 41 (when a step change occurs) the observation is again treated as an outlier with $p^{(112)}$ = .968 but as soon as the next observation becomes available the system can infer with considerable

$p^{(ijk)} \times 10^3$, m_t and b_t produced by 2PS (using the standard SSP) when applied to C8 Data

	(ijk)										2PS		MSM	
t	(111)	(121)	(131)	(141)	(211)	(311)	(411)	(112)	(113)	(114)	m_t	b_t	m_t	b_t
1	910	11	11	11	11	11	11	22	3	1	6.004	.000	6.004	.000
2	925	16	3	1	12	12	12	17	3	1	6.030	.009	6.030	.009
3	946	14	2	1	17	2	1	15	3	0	6.042	.010	6.042	.010
4	950	13	2	0	14	2	1	14	3	0	6.055	.011	6.055	.011
5	946	14	2	0	13	2	1	17	3	1	6.042	.004	6.042	.004
6	949	14	2	1	15	2	1	14	3	0	6.042	.003	6.042	.003
7	936	14	3	1	13	3	1	26	3	1	6.011	-.005	6.011	-.005
8	942	18	3	1	16	3	1	14	3	0	5.994	-.007	5.994	-.007
9	903	21	3	0	28	2	1	37	4	1	6.022	-.001	6.021	-.001
10	927	24	3	1	25	3	0	14	3	0	6.032	.001	6.033	.001
11	943	13	2	0	23	2	1	12	2	0	6.035	.001	6.035	.001
12	955	12	2	0	14	2	0	12	2	0	6.038	.001	6.038	.002
13	948	13	2	0	13	2	0	18	3	1	6.024	-.001	6.024	-.001
14	938	14	4	1	16	3	0	21	3	1	6.006	-.003	6.005	-.003
15	857	40	3	0	19	1	0	72	5	2	6.032	.001	6.032	.001
16	864	84	2	1	34	1	0	11	2	0	6.025	-.000	6.026	-.000
17	904	13	2	0	63	2	1	13	2	0	6.035	.001	6.036	.001
18	910	18	3	0	11	2	0	50	4	2	6.013	-.002	6.011	-.002
19	916	45	2	1	19	2	0	11	2	0	6.009	-.002	6.007	-.002
20	923	11	3	0	31	4	2	22	3	1	5.992	-.003	5.989	-.004
21	0	0	0	0	0	0	0	969	0	31	6.008	-.003	6.004	-.004
22	0	1000	0	0	0	0	0	0	0	0	5.992	-.003	5.990	-.003
23	0	0	0	0	1000	0	0	0	0	0	5.993	-.002	5.992	-.002
24	978	0	0	0	0	0	0	18	3	1	5.979	-.003	5.976	-.004
25	962	21	2	0	0	0	0	12	3	0	5.982	-.003	5.979	-.003
26	948	14	2	0	21	1	0	11	2	0	5.976	-.003	5.974	-.003
27	955	12	2	0	14	1	0	12	2	0	5.978	-.003	5.977	-.003
28	956	13	2	0	13	1	0	11	2	0	5.974	-.003	5.972	-.003
29	956	12	2	0	15	1	0	11	2	0	5.970	-.003	5.969	-.003
30	932	12	3	0	14	2	0	32	4	1	5.982	-.002	5.983	-.001
31	935	29	3	1	13	3	1	13	2	0	5.985	-.001	5.988	-.000
32	931	14	2	0	35	1	1	13	3	0	5.978	-.002	5.980	-.001
33	953	14	2	0	15	1	0	11	2	0	5.978	-.002	5.981	-.001
34	952	12	2	0	16	1	0	13	3	0	5.982	-.001	5.986	-.000
35	949	15	2	0	13	1	0	16	3	1	5.973	-.002	5.974	-.002
36	945	19	2	0	15	1	0	14	3	0	5.978	-.001	5.981	-.001
37	944	17	2	0	18	1	0	14	3	0	5.970	-.002	5.971	-.002
38	639	16	6	0	9	4	0	307	9	10	5.988	.000	5.994	.001
39	635	335	2	1	16	1	0	8	2	0	5.984	-.000	5.990	.000
40	503	8	8	1	75	28	15	344	8	11	6.014	.004	6.022	.005
41	1	0	0	0	0	0	0	968	0	31	6.016	.004	6.015	.004
42	12	355	2	627	3	0	0	0	0	0	5.827	-.002	5.899	.000
43	3	0	0	0	5	6	985	1	0	0	5.748	.002	5.750	.002
44	984	0	0	0	0	0	0	13	3	0	5.749	.002	5.744	.003
45	968	13	2	0	0	0	0	13	3	0	5.742	.000	5.739	.001
46	952	14	2	0	13	2	0	13	3	0	5.750	.002	5.749	.004
47	943	16	2	0	13	2	0	20	3	1	5.731	-.003	5.728	-.003
48	945	20	2	0	16	2	0	12	2	0	5.731	-.002	5.730	-.002
49	947	12	2	0	21	2	1	12	2	0	5.735	-.001	5.735	-.000
50	920	17	3	0	14	2	0	39	4	1	5.704	-.006	5.700	-.007

Table 8.4

51	929	33	2	1	17	2	0	12	2	0	5.697	-.006	5.693	-.007
52	888	13	3	0	57	2	0	31	4	1	5.716	-.002	5.715	-.002
53	899	55	2	0	12	1	0	26	3	1	5.691	-.005	5.689	-.006
54	891	20	3	1	68	2	0	13	2	0	5.677	-.006	5.673	-.007
55	948	12	2	0	18	2	1	13	3	0	5.662	-.007	5.658	-.008
56	929	18	2	0	15	1	0	29	4	1	5.675	-.005	5.673	-.005
57	932	23	3	1	21	2	0	14	3	0	5.678	-.004	5.678	-.004
58	918	12	5	1	17	7	2	34	4	1	5.694	-.002	5.696	-.001
59	917	47	2	0	13	1	0	16	3	1	5.679	-.004	5.681	-.003
60	904	14	2	1	61	1	0	14	3	0	5.666	-.005	5.667	-.004
61	927	19	2	0	19	1	0	27	3	1	5.679	-.003	5.682	-.002
62	918	19	5	2	26	3	0	23	3	1	5.692	-.001	5.696	.000
63	924	17	5	1	15	8	3	23	3	1	5.706	.001	5.712	.002
64	86	1	34	4	1	77	2	767	4	24	5.736	.009	5.734	.006
65	563	17	23	173	0	192	15	15	2	0	5.901	.045	5.866	.033
66	789	8	2	1	0	25	160	13	2	0	5.963	.050	5.952	.046
67	940	14	3	1	9	3	0	26	3	1	6.061	.067	6.050	.063
68	882	57	2	1	11	2	1	40	3	1	6.070	.049	6.065	.048
69	891	51	2	1	31	2	1	18	3	1	6.149	.057	6.146	.057
70	811	57	3	0	15	4	1	101	5	3	6.145	.043	6.141	.042
71	855	82	2	2	42	2	0	12	2	0	6.195	.045	6.192	.044
72	850	11	2	0	114	2	2	16	3	0	6.259	.048	6.258	.048
73	953	12	2	1	12	2	0	15	3	0	6.320	.050	6.320	.051
74	951	15	2	0	13	2	1	13	3	0	6.363	.049	6.364	.049
75	949	12	2	0	18	2	0	13	3	0	6.403	.048	6.403	.048
76	954	14	2	0	13	2	0	12	2	0	6.452	.048	6.452	.048
77	945	12	2	0	17	2	0	18	3	1	6.515	.050	6.516	.050
78	936	25	2	0	13	1	0	19	3	1	6.550	.048	6.550	.048
79	937	21	2	0	24	1	0	11	2	0	6.601	.048	6.602	.049
80	946	13	2	0	18	2	0	15	3	0	6.640	.047	6.639	.047
81	871	23	3	0	11	2	0	81	5	3	6.711	.050	6.713	.051
82	887	58	4	2	29	3	0	13	2	0	6.770	.051	6.773	.052
83	915	13	2	0	49	3	2	12	2	0	6.827	.051	6.831	.052
84	955	13	2	0	14	1	0	11	2	0	6.877	.051	6.880	.052
85	940	12	3	0	16	2	0	23	3	1	6.914	.050	6.915	.050
86	937	29	2	0	13	1	0	15	3	0	6.973	.051	6.975	.051
87	933	18	2	0	24	2	0	18	3	1	7.012	.050	7.013	.049
88	941	21	2	0	17	1	0	14	3	0	7.069	.050	7.071	.051
89	920	18	2	0	17	2	0	35	4	1	7.103	.049	7.102	.048
90	926	38	2	1	19	1	0	11	2	0	7.155	.049	7.153	.049
91	934	12	2	0	36	1	1	11	2	0	7.200	.049	7.198	.048
92	953	13	2	0	12	1	0	15	3	0	7.257	.049	7.257	.049
93	943	19	2	0	13	1	0	18	3	1	7.296	.048	7.294	.048
94	909	27	2	0	15	2	0	39	4	1	7.361	.050	7.361	.050
95	905	51	2	0	24	1	0	14	3	0	7.403	.049	7.402	.049
96	898	13	2	1	67	1	0	14	3	0	7.444	.048	7.441	.048
97	953	15	2	0	14	1	0	11	2	0	7.492	.049	7.490	.048
98	952	12	2	0	18	1	0	12	2	0	7.545	.049	7.544	.049
99	954	14	2	0	13	1	0	13	3	0	7.588	.048	7.586	.048
100	951	13	2	0	16	1	0	13	3	0	7.632	.048	7.628	.047

Table 8.4 cont.

confidence ($p^{(141)} = .627$) that the previous observation was not in fact an outlier but a step change. A further observation at time t = 43 is enough for the system to infer with maximum confidence ($p^{(411)} = .985$) that a step change did occur two periods back.

Finally starting from t = 61 (when a growth change takes place) the probabilities $p^{(131)}$ and $p^{(311)}$ increase considerably relative to their levels before t = 61, indicating that a growth change has been detected. It is important to note however that $p^{(311)}$ is much higher than $p^{(131)}$ and it is the contribution of the information associated with transition 311 that causes 2PS to respond faster to a growth change than the MSM. This is because the latter takes into account information associated with transition 31 (analogous to 131 in the 2PS) but inevitab ly ignores the more important information associated with transition 311, (as a result of the fact that its states are defined over a single point in time, in contrast to 2PS whose states are defined over two points in time). Comparing the growth responses of the two systems just after t = 61 (b_t columns in table 8.4), it can be seen clearly that the 2PS estimate of the growth, b_t, responds to the growth change faster than the equivalent estimate of the MSM.

In order to test the extent to which the 2PS responds consistently faster to growth changes than the MSM the following experiment has been performed. Consider 10 different realisations of an EWMA process (with $\sigma = \sqrt{V_\varepsilon} = 0.05$ like in C6 Data) and let us introduce a growth change of size $k_3\sigma$ at time $t = 21$ so that the true growth β_t is such that :

$$\begin{cases} \beta_t = 0 & t < 21 \\ \beta_t = k_3\sigma & t \geqslant 21 \end{cases}$$

Three sizes of growth change are considered :

	(i)	$k_3 = \frac{1}{2}$	i.e. $\frac{1}{2}\sigma$	growth change at t = 21
	(ii)	$k_3 = 1$	i.e. 1σ	" " " "
and	(iii)	$k_3 = 1.5$	i.e. 1.5σ	" " " "

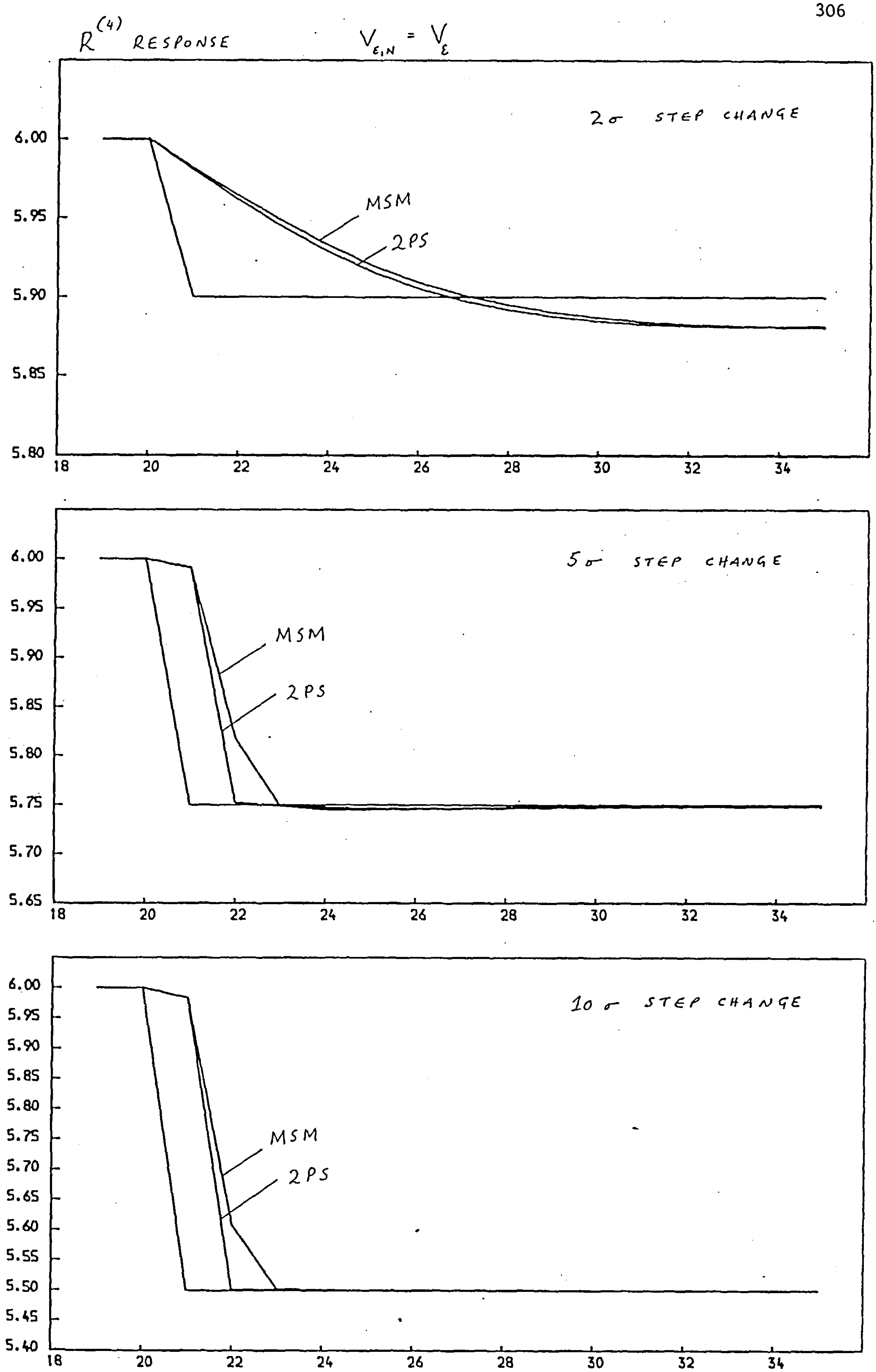
The growth estimates b_t produced by the MSM and 2PS and corresponding to $k_3 = \frac{1}{2}$, 1, 1.5 are shown in Appendix M, tables M.1a, M.1b, M.1c and M.2a, M2b, M.2c respectively for t = 1, 2, ...50. From these tables, it can be seen that although there is little difference between the MSM and 2PS system growth responses for the $\frac{1}{2}\sigma$ growth changes, the latter system consistently responds faster than the MSM to more abrupt growth changes of order 1σ or larger. This can also be seen from the $R^{(3)}$ response of figure 8.6a. The other performance measures $R^{(1)}$, $R^{(2)}$, $R^{(4)}$ (as defined in chapter 4) are given together with those of the MSM in table 8.5 and figure 8.6b respectively.

Table 8.5

	$R^{(1)}$ MSE	$R^{(2)}$ z response to outliers of size: 4σ	10σ	20σ
MSM	100	.3	.3	.7
2PS	96	.3	.3	.7

$R^{(3)}$ RESPONSE
$V_{\varepsilon,N} = V_{\varepsilon}$
0.050
0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
-0.005
18
20
22
24
26
28
30
32
34
$\frac{1}{2}\sigma$ GROWTH CHANGE
2PS
MSM
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
1σ GROWTH CHANGE
2PS
MSM
0.10
0.08
0.06
0.04
0.02
0.00
$1{\cdot}5\,\sigma$ GROWTH CHANGE
2PS
MSM

FIGURE 8.6a



$R^{(4)}$ RESPONSE
$V_{E,N} = V_E$
2σ STEP CHANGE
MSM
2PS
6.00
5.95
5.90
5.85
5.80
18
20
22
24
26
28
30
32
34
5σ STEP CHANGE
MSM
2PS
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
10σ STEP CHANGE
MSM
2PS
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
5.60
5.55
5.50
5.45
5.40

FIGURE 8.6b

The results of table 8.5 imply that over a quiet period when no abrupt changes take place the 2PS has a MSE which is 4% smaller than that of the MSM. This can be explained with reference to tables 8.4 and 8.2. An important factor affecting the stability of the system and consequently the MSE value, is the transition probability $p^{(31)}$ in the case of the MSM and $p^{(311)}$ in the case of the 2PS. The higher these probabilities the more responsive (and the higher the MSE) the system becomes, since more weight is placed on the transitions $3 \to 1$ and $31 \to 11$ (for MSM and 2PS respectively) which imply that a growth change has taken place. Comparing $p^{(31)}$ from tables 8.2 with $p^{(311)}$ from table 8.4 before $t = 41$ when no genuine growth change takes place, it can be seen that the latter transition probability is always less than or equal to $p^{(31)}$ which explains why the 2PS is more stable over quiet periods and has therefore a smaller MSE value. However, just after $t = 61$ (when a growth change takes place) $p^{(311)}$ of 2PS is higher than $p^{(31)}$ of MSM which also explains why 2PS responds faster to growth changes than MSM as illustrated in figure 8.6a.

In addition from table 8.5 and figure 8.6b we can conclude that the 2PS responds to outliers and step changes in a nearly identical way to the MSM except that the 2PS (like the 1PS) responds to step changes marginally faster.

Finally the 2PS is compared with MSM on its robustness to errors in the nominated variance $V_{\varepsilon,N}$. Table 8.6 shows $R^{(1)}$ and $R^{(2)}$. $R^{(3)}$ and $R^{(4)}$ are illustrated in figures 8.7a, 8.7b and 8.8a, 8.8b corresponding to $V_{\varepsilon,N}$ underestimating and overestimating V_ε by a factor of 2 respectively. Larger errors in $V_{\varepsilon,N}$ are not considered since from chapter 6 we have seen that the on line variance estimation methods produce estimates which are never in error from V_ε by more than a factor of two.

TABLE 8.6.

		$R^{(1)}$ MSE	$R^{(2)}$ z response to outliers of size: 4σ	10σ	20σ
MSM	$V_{\varepsilon,N} = \frac{1}{2} V_\varepsilon$	102	.1	.3	.7
	$V_{\varepsilon,N} = 2 V_\varepsilon$	99	.8	.3	.7
2PS	$V_{\varepsilon,N} = \frac{1}{2} V_e$	99	.1	.3	.7
	$V_{\varepsilon,N} = 2 V_\varepsilon$	96	.7	.3	.7

It can be seen that even when our nominated variance is in error from V_ε, the 2PS has a lower MSE than the MSM, as well as a faster response to growth and step changes. The response to outliers is not significantly different.

In conclusion it can be said that the 2PS is more robust (to errors in V_ε) yet more responsive to genuine changes than the MSM. In addition by reducing the allowed transitions (as in 1PS) the computation time is approximately ½ of that required by the MSM.


R(3) RESPONSE
V_{E,N} = ½ V_E
½ σ GROWTH CHANGE
2PS
MSM
1σ GROWTH CHANGE
2PS
MSM
1·5 σ GROWTH CHANGE
2PS
MSM
0.050
0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
-0.005
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0.10
0.08
0.06
0.04
0.02
0.00
18
20
22
24
26
28
30
32
34

FIGURE 8.7

R(4) RESPONSE
V E,N = 1/2 V E
2 σ STEP CHANGE
5 σ STEP CHANGE
10 σ STEP CHANGE
MSM
2PS
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
5.60
5.55
5.50
5.45
5.40
18
20
22
24
26
28
30
32
34

FIGURE 8.7L

$R^{(3)}$ RESPONSE $\qquad V_{\varepsilon,N} = 2\,V_{\varepsilon}$

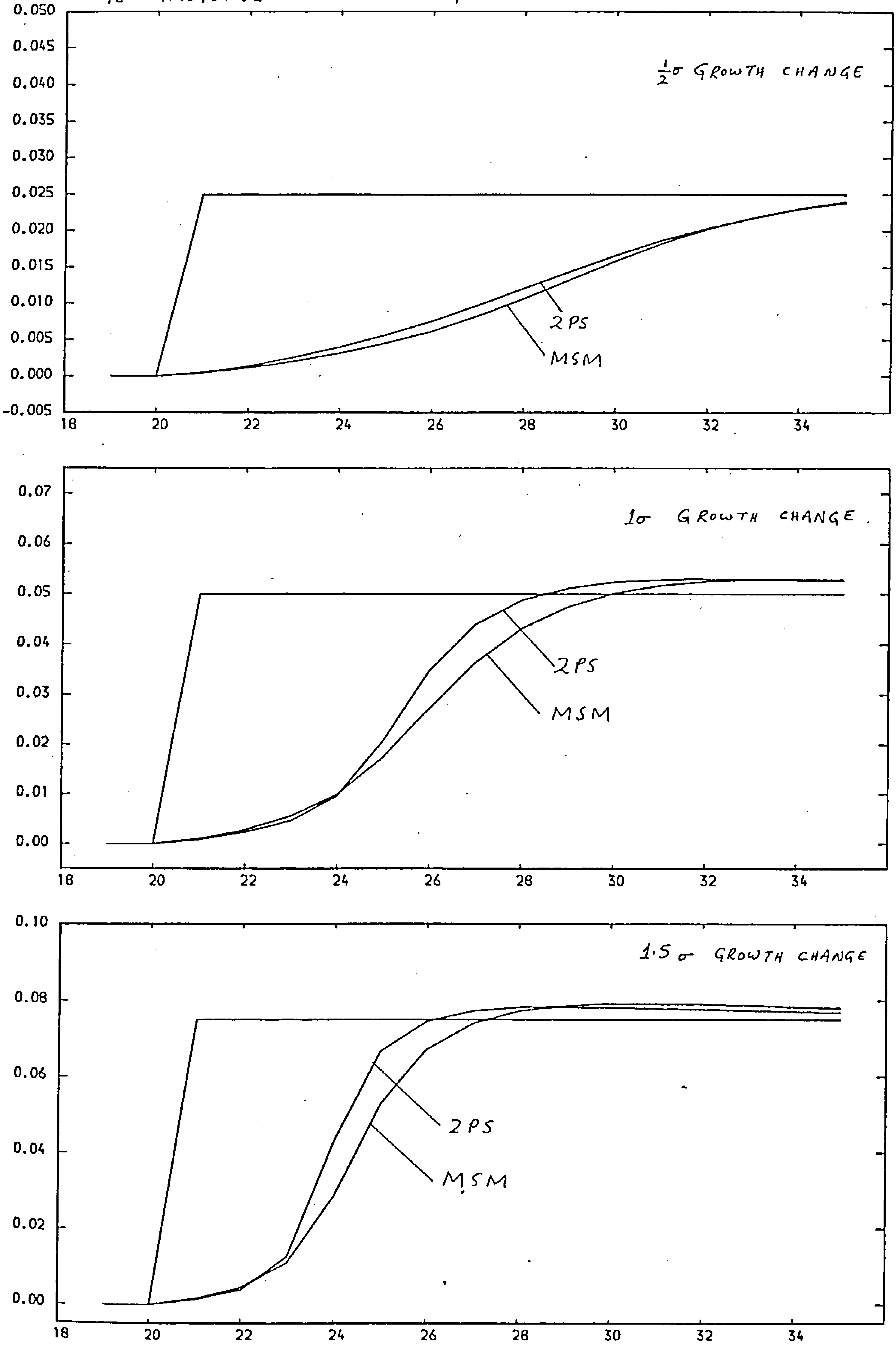


FIGURE 9.8

$R^{(4)}$ RESPONSE
$V_{E,N} = 2\, V_E$
2σ STEP CHANGE
2PS
MSM
6.00
5.95
5.90
5.85
5.80
18
20
22
24
26
28
30
32
34
5σ STEP CHANGE
2PS
MSM
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
18
20
22
24
26
28
30
32
34
10σ STEP CHANGE
2PS
MSM
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
5.60
5.55
5.50
5.45
5.40
18
20
22
24
26
28
30
32
34

8.5. The "three-three" System (33S)

The formulation of the MSM described in chapter 3 assumes that the transition probabilities $\{\Pi^{(1)}, \Pi^{(2)}, \Pi^{(3)}, \Pi^{(4)}\}$ are time invariant and further depend only on the state the process moves to at any time t and independent of the previous state at time t-1. This can be viewed as a simplification of a more general set up where the probability of transition $i \to j$ in the time interval $(t - 1, t)$ is described by a square matrix $[\Pi^{(ij)}]$.

If $\Pi^{(ij)} = \Pi^{(j)}$ for $i, j = 1, 2, 3, 4$ then we obtain the four MSM transition probabilities. Using this formulation it was shown (in chapter 5) that a faster growth response could be achieved by changing the balance of the Π's in the standard SSP and in particular by increasing $\Pi^{(3)}$. However, this improvement was offset by a large decrease in stability during quiet periods. This result could be expected from the following argument. Consider the process at time t and recall (from section 3.3) that the posterior probabilities corresponding to a state transition $i \to 3$ can be written as follows:

$$p^{(i3)} \propto L^{(i3)} \cdot p_{t-1}^{(i)} \cdot \Pi^{(3)} \qquad (8.5.1)$$

where

(i) $L^{(i3)}$ is the likelihood of the latest observation y_t conditional on an $i \to 3$ transition from t-1 to t.

(ii) $p^{(i)}$ is the probability (posterior to y_{t-1}) that the process was in state i at time t-1.

and (iii) $\Pi^{(3)}$ is the constant probability of occurrence of state 3 (i.e. growth change) at any time t independent of the state of the process at t - 1.

From (8.5.1) above it can be seen that increasing $\Pi^{(3)}$ results in higher $p^{(i3)}$ and therefore the probability (posterior to y_t) of a growth change $p_t^{(3)}$ also increases since:

$$p_t^{(3)} = \sum_i p^{(i3)} \text{ for } i = 1, 2, 3, 4. \qquad (8.5.2)$$

The higher probabilities $p_t^{(3)}$ achieved in this way imply that more weight is given to state 3 which in turn causes the system to be more responsive at all times. Very crudely we can say that increasing $\Pi^{(3)}$ in the MSM has an analogous effect to increasing the smoothing constant α in an EWMA model. Ideally we would like to operate with a low α during quiet periods and a high α at points of "change". In the MSM something roughly analogous would be to operate with a low $p_t^{(3)}$ during quiet periods and a high $p_t^{(3)}$ just after a growth change has taken place (This idea is similar to that proposed by Trigg and Leach [47] whereby the response rate in exponential smoothing is varied according to the extent to which biased forecasts are being obtained). One possible approach is to drop the assumption of $\Pi^{(ij)} = \Pi^{(j)}$ and all i and in particular for transitions from state 3. The 33S assumes a lower $\Pi^{(31)}$ than $\Pi^{(1)}$ and a higher $\Pi^{(33)}$ than $\Pi^{(3)}$. Because of this high probability associated with transition 3 → 3 it has been called the "three-three" system. The $[\Pi^{(ij)}]$ transition probability matrix used by 33S is shown below together with that implicitly used by the MSM:

$$\begin{array}{cc} & \text{MSM} \\ i \nearrow j & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{bmatrix} \Pi^{(1)} & \Pi^{(2)} & \Pi^{(3)} & \Pi^{(4)} \\ " & " & " & " \\ " & " & " & " \\ " & " & " & " \end{bmatrix} \end{array} \qquad \begin{array}{cc} & \text{33S} \\ i \nearrow j & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{bmatrix} \Pi^{(1)} & \Pi^{(2)} & \Pi^{(3)} & \Pi^{(4)} \\ " & " & " & " \\ \Pi^{(31)} & " & \Pi^{(33)} & " \\ \Pi^{(1)} & " & \Pi^{(3)} & " \end{bmatrix} \end{array}$$

where $\Pi^{(31)} < \Pi^{(1)}$ and $\Pi^{(33)} > \Pi^{(3)}$

A reasonable choice for $\Pi^{(33)}$ (as will be shown in the next section) is .4 and this together with the standard SSP values for $\Pi^{(j)}$ $j = 1, 2, 3, 4$, gives the following 33S transition matrix.

$$\begin{array}{cc} i \nearrow j & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{bmatrix} .900 & .094 & .003 & .003 \\ .900 & .094 & .003 & .003 \\ .503 & .094 & .400 & .003 \\ .900 & .094 & .003 & .003 \end{bmatrix} \end{array}$$

The effect of the choice of $\Pi^{(33)}$ as well as various other comparisons between the MSM, 2PS and 33S are examined in the next section.

8.5.1. Numerical illustration of 33S

The response of the 33S to C8 Data will first be shown for different values of $\Pi^{(33)}$. Note that $\Pi^{(33)}$ is the only extra parameter (additional to those in an SSP used by the MSM) introduced by the 33S, since given a value for $\Pi^{(33)}$ then all the remaining probabilities in the third row of the transition matrix can be expressed in terms of $\Pi^{(33)}$ and $\Pi^{(j)}$ from the SSP:

$$\left.\begin{aligned} \Pi^{(34)} &= \Pi^{(4)} \\ \Pi^{(32)} &= \Pi^{(2)} \\ \text{and} \qquad \Pi^{(31)} &= 1 - \Pi^{(4)} - \Pi^{(2)} - \Pi^{(33)} \end{aligned}\right\} \qquad (8.5.1.1)$$

Clearly if $\Pi^{(33)} = \Pi^{(3)}$ then the 33S reduces to the MSM.

In tables 8.7 and N.1, N.2, N.3 (see Appendix N) we tabulate:

(i) probabilities $p^{(ij)}$ produced by 33S

(ii) system growth b_t produced by 33S

corresponding to $\Pi^{(33)}$ = .4 and .2, .6, .8 respectively. Comparing these four tables with table 8.2 earlier (showing $\tilde{p}^{(ij)}$ produced by MSM) a number of properties of 33S can be illustrated :

TABLE 8.7

	$\rho^{(11)}$	$\rho^{(21)}$	$\rho^{(31)}$	$\rho^{(41)}$	$\rho^{(12)}$	$\rho^{(22)}$	$\rho^{(32)}$	$\rho^{(42)}$	$\rho^{(13)}$	$\rho^{(23)}$	$\rho^{(33)}$	$\rho^{(43)}$	$\rho^{(14)}$	$\rho^{(24)}$	$\rho^{(34)}$	$\rho^{(44)}$	33S b_t	MSM b_t
1	876	91	2	3	21	2	0	0	3	0	1	0	1	0	0	0	.000	.000
2	959	15	2	1	18	0	0	0	3	0	1	0	1	0	0	0	.009	.009
3	965	13	2	1	15	0	0	0	3	0	1	0	0	0	0	0	.010	.010
4	966	12	2	0	15	0	0	0	3	0	1	0	0	0	0	0	.011	.011
5	962	13	2	0	18	0	0	0	3	0	1	0	1	0	0	0	.004	.004
6	966	13	2	1	14	0	0	0	3	0	1	0	0	0	0	0	.003	.003
7	951	13	2	0	27	0	0	0	3	0	2	0	1	0	0	0	-.005	-.005
8	961	16	2	1	14	0	0	0	3	0	1	0	0	0	0	0	-.007	-.007
9	932	20	2	0	38	1	0	0	4	0	1	0	1	0	0	0	-.001	-.001
10	955	23	2	1	14	1	0	0	3	0	1	0	0	0	0	0	.001	.001
11	969	13	1	0	12	0	0	0	2	0	1	0	0	0	0	0	.001	.001
12	971	11	1	0	12	0	0	0	2	0	1	0	0	0	0	0	.001	.002
13	963	12	2	0	18	0	0	0	3	0	1	0	1	0	0	0	-.001	-.001
14	956	13	3	1	21	0	0	0	3	0	2	0	1	0	0	0	-.003	-.003
15	875	39	1	0	74	2	0	0	5	0	1	0	2	0	0	0	.001	.001
16	902	79	1	1	12	1	0	0	2	0	1	0	0	0	0	0	-.000	-.000
17	966	13	1	0	14	0	0	0	3	0	1	0	0	0	0	0	.001	.001
18	918	20	2	0	51	1	0	0	4	0	2	0	2	0	0	0	-.002	-.002
19	940	41	1	1	11	1	0	0	2	0	1	0	0	0	0	0	-.002	-.002
20	958	12	2	0	22	0	0	0	3	0	2	0	1	0	0	0	-.004	-.004
21	0	0	0	0	942	23	4	1	0	0	0	0	30	1	0	0	-.004	-.004
22	0	984	0	0	0	13	0	0	0	3	0	0	0	0	0	0	-.003	-.003
23	975	8	0	0	12	0	0	0	3	0	0	0	0	0	0	0	-.003	-.002
24	963	13	2	0	18	0	0	0	3	0	1	0	1	0	0	0	-.004	-.004
25	964	19	0	0	13	0	0	0	3	0	0	0	0	0	0	0	-.003	-.003
26	970	13	1	0	12	0	0	0	2	0	1	0	0	0	0	0	-.003	-.003
27	970	12	1	0	13	0	0	0	3	0	1	0	0	0	0	0	-.002	-.003
28	970	13	1	0	12	0	0	0	2	0	1	0	0	0	0	0	-.003	-.003
29	972	11	1	0	12	0	0	0	2	0	1	0	0	0	0	0	-.003	-.003
30	947	12	2	0	32	0	0	0	4	0	2	0	1	0	0	0	-.000	-.001
31	955	25	2	1	13	0	0	0	3	0	1	0	0	0	0	0	.000	-.000
32	966	13	1	0	15	0	0	0	3	0	1	0	0	0	0	0	-.001	-.001
33	969	15	1	0	12	0	0	0	2	0	1	0	0	0	0	0	-.001	-.001
34	971	11	1	0	13	0	0	0	3	0	1	0	0	0	0	0	.000	-.000
35	961	14	1	0	19	0	0	0	3	0	1	0	1	0	0	0	-.001	-.002
36	959	21	1	0	14	0	0	0	3	0	1	0	0	0	0	0	-.000	-.001
37	962	16	1	0	16	0	0	0	3	0	1	0	1	0	0	0	-.002	-.002
38	704	22	4	0	245	4	1	0	8	0	4	0	8	0	0	0	.002	.001
39	716	265	1	1	9	3	0	0	2	1	1	0	0	0	0	0	.000	.000
40	654	23	9	0	282	4	1	0	8	0	8	0	9	0	0	0	.005	.005
41	0	1	0	0	657	291	13	7	0	0	0	0	21	9	0	0	.005	.004
42	10	262	0	359	0	340	0	12	0	3	0	1	0	11	0	0	-.000	.000
43	890	4	11	14	45	21	0	1	3	0	8	0	1	1	0	0	.002	.002
44	958	20	2	2	13	1	0	0	3	0	1	0	0	0	0	0	.003	.003
45	969	11	2	0	14	0	0	0	3	0	1	0	0	0	0	0	.001	.001
46	967	13	1	0	14	0	0	0	3	0	1	0	0	0	0	0	.004	.004
47	957	16	2	0	21	0	0	0	3	0	1	0	1	0	0	0	-.003	-.003
48	962	18	1	1	13	0	0	0	3	0	1	0	0	0	0	0	-.002	-.002
49	969	12	1	0	13	0	0	0	3	0	1	0	0	0	0	0	-.000	-.000
50	935	17	2	0	38	1	0	0	4	0	2	0	1	0	0	0	-.007	-.007

$\rho^{(ij)} \times 10^3$ and b_t (estimate of β_t) produced by 33S when applied to CB Data with $\pi^{(33)} = .4$ and using the standard SSP

The last column is the MSM b_t for the same data.

TABLE 8.7 cont.

											$\rho^{(33)}$							
51	951	29	2	1	12	0	0	0	2	0	1	0	0	0	0	0	-.007	-.007
52	942	14	2	0	35	0	0	0	4	0	2	0	1	0	0	0	-.002	-.002
53	908	58	1	0	26	1	0	0	3	0	1	0	1	0	0	0	-.006	-.006
54	961	19	2	1	13	0	0	0	3	0	1	0	0	0	0	0	-.007	-.007
55	968	12	2	0	13	0	0	0	3	0	1	0	0	0	0	0	-.008	-.008
56	942	17	2	0	33	0	0	0	4	0	1	0	1	0	0	0	-.005	-.005
57	955	22	2	1	14	0	0	0	3	0	1	0	0	0	0	0	-.004	-.004
58	941	12	4	1	34	1	0	0	4	0	3	0	1	0	0	0	-.001	-.001
59	932	44	1	0	17	1	0	0	3	0	1	0	1	0	0	0	-.003	-.003
60	964	14	2	1	15	0	0	0	3	0	1	0	0	0	0	0	-.005	-.004
61	947	19	1	0	27	0	0	0	3	0	1	0	1	0	0	0	-.002	-.002
62	949	17	3	1	22	1	0	0	3	0	2	0	1	0	0	0	-.000	.000
63	952	16	3	1	21	1	0	0	3	0	2	0	1	0	0	0	.002	.002
64	126	2	51	3	724	16	4	1	5	0	44	0	23	1	0	0	.010	.006
65	541	30	107	150	16	64	4	2	2	1	79	0	1	2	0	0	.039	.033
66	875	16	49	3	17	2	2	0	3	0	34	0	1	0	0	0	.048	.046
67	910	13	25	1	27	1	1	0	3	0	18	0	1	0	0	0	.063	.063
68	891	53	7	1	35	1	1	0	3	0	6	0	1	0	0	0	.048	.048
69	924	45	3	1	20	1	0	0	3	0	2	0	1	0	0	0	.056	.057
70	827	59	4	0	95	2	1	0	5	0	4	0	3	0	0	0	.042	.042
71	900	77	2	2	14	1	0	0	2	0	1	0	0	0	0	0	.044	.044
72	963	12	2	1	18	0	0	0	3	0	1	0	1	0	0	0	.048	.048
73	965	13	2	1	15	0	0	0	3	0	1	0	0	0	0	0	.050	.051
74	967	14	1	0	13	0	0	0	3	0	1	0	0	0	0	0	.049	.049
75	969	12	1	0	14	0	0	0	3	0	1	0	0	0	0	0	.048	.048
76	970	13	1	0	12	0	0	0	2	0	1	0	0	0	0	0	.048	.048
77	963	12	1	0	19	0	0	0	3	0	1	0	1	0	0	0	.050	.050
78	951	24	1	0	19	0	0	0	3	0	1	0	1	0	0	0	.048	.048
79	963	19	1	0	12	0	0	0	2	0	1	0	0	0	0	0	.048	.049
80	966	12	1	0	15	0	0	0	3	0	1	0	0	0	0	0	.047	.047
81	879	23	2	0	83	1	0	0	5	0	2	0	3	0	0	0	.050	.051
82	922	53	3	2	13	1	0	0	3	0	2	0	0	0	0	0	.051	.052
83	968	13	2	0	13	0	0	0	3	0	1	0	0	0	0	0	.052	.052
84	970	13	1	0	12	0	0	0	2	0	1	0	0	0	0	0	.052	.052
85	956	11	2	0	24	0	0	0	3	0	1	0	1	0	0	0	.050	.050
86	951	28	1	0	15	0	0	0	3	0	1	0	0	0	0	0	.051	.051
87	957	18	1	0	19	0	0	0	3	0	1	0	1	0	0	0	.049	.049
88	959	21	1	0	14	0	0	0	3	0	1	0	0	0	0	0	.050	.051
89	937	18	2	0	35	1	0	0	4	0	1	0	1	0	0	0	.048	.048
90	947	35	1	1	12	0	0	0	2	0	1	0	0	0	0	0	.049	.049
91	971	12	1	0	12	0	0	0	2	0	1	0	0	0	0	0	.048	.048
92	966	12	1	0	16	0	0	0	3	0	1	0	1	0	0	0	.049	.049
93	957	19	1	0	18	0	0	0	3	0	1	0	1	0	0	0	.048	.048
94	921	26	2	0	42	1	0	0	4	0	1	0	1	0	0	0	.050	.050
95	927	52	1	0	15	1	0	0	3	0	1	0	0	0	0	0	.049	.049
96	965	13	2	1	15	0	0	0	3	0	1	0	0	0	0	0	.048	.048
97	969	15	1	0	11	0	0	0	2	0	1	0	0	0	0	0	.048	.048
98	971	11	1	0	13	0	0	0	3	0	1	0	0	0	0	0	.049	.049
99	968	13	1	0	13	0	0	0	3	0	1	0	0	0	0	0	.048	.048
100	969	12	1	0	13	0	0	0	3	0	1	0	0	0	0	0	.047	.047

(a) The response of 33S is almost identical to that of the MSM for periods of no change, outliers and step changes and this is true independent of the value of $\Pi^{(33)}$.

This is clearly shown by comparing table 8.2 with tables 8.7, N.1, N.2, N.3 around t = 21 and t = 41 (when an outlier and a step change occur respectively) as well as during t = 1 to 40 which is a period of no change (apart from the outlier at t = 21).

(b) Following a genuine growth change (as at t = 61) $p^{(33)}$ and as a result $p_t^{(3)}$ in the 33S, becomes significantly higher than in the MSM and consequently 33S produces a faster response to growth changes than the MSM. This is also clear by comparison of the system estimates of the growth, b_t, as given in the last two columns of tables 8.7, N.1, N.2, N.3.

(c) $p^{(31)}$ which through the *collapsing process* feeds information from state 3 (growth change) to state 1 (no change) is consistently smaller in the 33S at all times except during a period following a growth change. It is therefore expected that because 33S gives less weight to the growth change state, it will be more stable and will therefore have a smaller MSE than the MSM at all times except following a growth change period. This can be confirmed in terms of the $R^{(1)}$ measure as defined in chapter 4 and of course employing the data of that chapter. $R^{(1)}$ together with $R^{(2)}$ which describes the response of the system to outliers is shown below in table 8.8.:

Table 8.8

		$R^{(1)}$ MSE	$R^{(2)}$ z response to outliers of size: 4 σ	10 σ	20 σ
MSM		100	.3	.3	.7
33S	$\Pi^{(33)}$ = .2	100	.3	.3	.7
	" = .4	99	.3	.3	.7
	" = .6	98	.3	.3	.7
	" = .8	97	.4	.3	.7

The system growth and level responses $R^{(3)}$, $R^{(4)}$ corresponding to $\Pi^{(33)}$ = .4 (which will be recommended for reasons to be explained) are graphed in figures 8.9a and 8.9b respectively together with the equivalent MSM responses. It can be seen that $R^{(4)}$ is identical in both 33S and MSM which confirms the point made in (a) above.
The same conclusion is valid for all other $\Pi^{(33)}$ values. The 33S has however a faster $R^{(3)}$ response than MSM with the speed of response increasing as $\Pi^{(33)}$ increases. Comparing figures 8.9a and 8.6a it can be seen that although the 33S with $\Pi^{(33)}$ = .4 responds faster than the MSM to growth changes its speed of response is marginally lower than that of 2PS. However, higher values of $\Pi^{(33)}$ (say of the order of .6 or larger) produce even faster response than the 2PS.


R(3) RESPONSE
π(33) = ·4
V ε,N = V ε
1/2 σ GROWTH CHANGE
33S
MSM
1σ GROWTH CHANGE
33S
MSM
1·5 σ GROWTH CHANGE
33S
MSM

FIGURE 8.9a

$R^{(4)}$ RESPONSE
$\pi^{(33)} = .4$
$V_{\varepsilon,N} = V_{\varepsilon}$
2 σ STEP CHANGE
MSM
33S
6.00
5.95
5.90
5.85
5.80
18
20
22
24
26
28
30
32
34
5 σ STEP CHANGE
MSM
33S
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
18
20
22
24
26
28
30
32
34
10 σ STEP CHANGE
MSM
33S
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
5.60
5.55
5.50
5.45
5.40
18
20
22
24
26
28
30
32
34

FIGURE 8.9b

(d) As we have already seen, increasing $\Pi^{(33)}$ causes the transition probabilities $p^{(33)}$ just after a growth change to increase significantly and remain high for some time. It is convenient at this point to define T as the number of time periods following a growth change during which $p^{(33)}$ is unusually high. There is a penalty of instability associated with high T because as long as $p^{(33)}$ remains unusually high, relatively more weight is given to the *growth change state* (and therefore the most recent observations). As a result the growth estimates b_t have a higher variability around the true β_t level and there is a risk of substantial error being introduced in b_t if significantly large forecast errors occur during the period of unusually high $p^{(33)}$. Clearly this penalty and risk increase with T which in turn increases with $\Pi^{(33)}$. Hence a choice of $\Pi^{(33)}$ has to be made subjectively according to the particular application for which the system is being used. These effects are to some extent noticeable in tables N.1. 8.7 , N.2, N.3 where T is approximately 4, 7, 10 and 16 when $\Pi^{(33)}$ = .2, .4, .6 and .8 respectively. However a better understanding can be achieved by examining the b_t estimates and average T ($\bar{T}$) values across the 10 realisations used in section 8.4.2 to produce the tables in Appendix M. Recall that each realisation exhibits a growth change of size $k_3\sigma$ at t=21 and the three usual values of k_3 = ½, 1, 1.5 have been considered.

Given a value for $\Pi^{(33)}$, the average T values (across the ten realisations) for different k_3 are very similar and are tabulated in table 8.9. The growth estimates produced by 33S with $\Pi^{(33)} = .2, .4, .6, .8$ and for $k_3 = \frac{1}{2}, 1, 1.5$ are shown in tables P.1., P.2., ...P.12 of Appendix P, illustrating the extent to which higher $\Pi^{(33)}$ lead to b_t instability.

Table 8.9

$\Pi^{(33)}$	.2	.4	.6	.8
Average T ($\overline{T}$)	5	8	13	19

An interesting observation can be made with reference to table P.2 (33S with $\Pi^{(33)} = .4$) and table M.2a (2PS) in Appendices P and M respectively. Consider the b_t estimates after t = 21 (when a growth change occurs and β_t jumps from zero to a new level of $\frac{1}{2}\sigma = .025$) for realisation 1. At t = 24 an unintentional discontinuity (due to a very high noise contribution) combined with the growth change which started 3 periods earlier cause both systems to react abruptly with the b_t estimates of 2PS and 33S going up to 0.040 and 0.036 at t = 24 from 0.001 and 0.002 at t = 23 respectively. In the case of 33S, this pushes the $p^{(33)}$ probability at a high level and introduces a bias in b_t. This can be seen clearly by comparing the b_t estimates of 33S and 2PS near time t = 30. The 2PS estimates are fairly stable around a level of 0.030 (recall that the true level is $\beta_t = 0.025$) while the 33S estimates are more unstable and over estimate β_t much more, fluctuating around a level of .038.

This example illustrates the point made earlier namely that the 33S runs the risk of instability and bias in b_t if a discontinuity occurs just after a growth change. This risk increases with $\Pi^{(33)}$ since as can be seen from table 8.9 $\overline{T}$ is an increasing function of $\Pi^{(33)}$. For a value of $\Pi^{(33)} = .4$ for example 33S is at risk (on average) for 8 time periods following a growth change while for $\Pi^{(33)} = .8$ it is at risk for 19 periods.

For most forecasting applications and particularly for medium term forecasts a value of T greater than 10 is unacceptable so that our choice for $\Pi^{(33)}$ is restricted to a value of .4 or smaller. Using $\Pi^{(33)} = 0.4$ the 33S is then comparable to 2PS. The latter system has a smaller MSE (see $R^{(1)}$ in tables 8.8 and 8.5) and a faster growth response (see figures 8.6a and 8.9a). The response to outliers and step changes ($R^{(2)}$ and $R^{(4)}$) are almost identical but the 2PS does not suffer from the risk of bias and instability present in the 33S after a growth change. It can be concluded that the 2PS is a better forecasting system than the 33S.

Finally, for completeness, the recommended form of 33S is compared with the MSM on its robustness to errors in the nominated variance $V_{\varepsilon,N}$. Table 8.10 shows $R^{(1)}$ and $R^{(2)}$. $R^{(3)}$ and $R^{(4)}$ are illustrated in figures 8.10a, 8.10b, and 8.11a, 8.11b corresponding to $V_{\varepsilon,N}$ underestimating and overestimating V_ε by a factor of 2 respectively.

Table 8.10

		$R^{(1)}$ MSE	$R^{(2)}$ z response to outliers of size: 4σ	10σ	20σ
MSM	$V_{\varepsilon,N} = \frac{1}{2} V_{\varepsilon}$	102	.1	.3	.7
	" $= 2 V_{\varepsilon}$	99	.8	.3	.7
33S	$V_{\varepsilon,N} = \frac{1}{2} \alpha v_{\varepsilon}$	102	.1	.3	.7
$\Pi^{(33)} = .4$	" $= 2 \alpha V_{\varepsilon}$	98	.8	.3	.7

It can be seen that when our nominated variance $V_{\varepsilon,N}$ is in error from the true variance V_{ε}, 33S and MSM produce practically identical responses to outliers, step changes and periods of no change. 33S's speed of response to growth changes is however faster as we have already seen for the case of $V_{\varepsilon,N} = V_{\varepsilon}$.

The 33S was an obvious line of development but has proved inferior to 2PS which should be preferred in forecasting situations. However the 33S might have a use in a monitoring role. If it is required only to detect growth changes then the 33S with a very high value of $\Pi^{(33)}$, say .8, would respond faster than the MSM (and even the 2PS). In addition it would be more stable in periods of no change as implied by the MSE values in table 8.8 corresponding to the MSM and the 33S with $\Pi^{(33)} = .8$.


$R^{(3)}$ RESPONSE
$\pi^{(33)} = .4$
$V_{\varepsilon,N} = \frac{1}{2} V_\varepsilon$
$\frac{1}{2}\sigma$ GROWTH CHANGE
33S
MSM
0.050
0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
-0.005
18
20
22
24
26
28
30
32
34
1σ GROWTH CHANGE
33S
MSM
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
1.5σ GROWTH CHANGE
33S
MSM
0.10
0.08
0.06
0.04
0.02
0.00

FIGURE 8.10a


$R^{(4)}$ RESPONSE
$\pi^{(33)} = \cdot 4$
$V_{\varepsilon,N} = \frac{1}{2} V_{\varepsilon}$
2σ STEP CHANGE
335
MSM
6.00
5.95
5.90
5.85
5.80
18
20
22
24
26
28
30
32
34
5σ STEP CHANGE
335
MSM
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
10σ STEP CHANGE
335
MSM
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
5.60
5.55
5.50
5.45
5.40

FIGURE 8.10b

$R^{(3)}$ RESPONSE
$\pi^{(33)} = .4$
$V_{\varepsilon,N} = 2V_\varepsilon$
$\frac{1}{2}\sigma$ GROWTH CHANGE
33S
MSM
0.050
0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
-0.005
18
20
22
24
26
28
30
32
34
1σ GROWTH CHANGE
33S
MSM
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
1.5σ GROWTH CHANGE
33S
MSM
0.10
0.08
0.06
0.04
0.02
0.00

FIGURE 8.11a

$R^{(4)}$ RESPONSE $\pi^{(33)} = .4$ $V_{\varepsilon,N} = 2V_{\varepsilon}$
2σ STEP CHANGE
335
MSM
6.00
5.95
5.90
5.85
5.80
18
20
22
24
26
28
30
32
34
5σ STEP CHANGE
335
MSM
6.00
5.95
5.90
5.85
5.80
5.75
5.70
5.65
10σ STEP CHANGE
335
MSM
5.60
5.55
5.50
5.45
5.40

FIGURE 8.11b

8.6. Concluding remarks

Three alternative formulations of the MSM have been described and illustrated.

The 1PS excludes a number of unlikely state transitions thus reducing the computational effort required by the MSM by a factor of two. Its response to the different states considered, is practically identical to that of the MSM with the exception of medium sized (or larger) step changes which are recognised faster by the 1PS.

Similar cost savings are achieved by the 2PS which in line with the 1PS considers only a small number of state transitions excluding the rest as unlikely. There is one fundamental difference between the 2PS and the MSM in the definition of a *state*. The former defines a *state* over two consecutive points in time and consequently calculates and assesses likelihoods based on 3 observations instead of 2 as in the MSM. The 2PS responds to outliers in a practically identical way to the MSM but produces a faster response (like the 1PS) to step changes. Over a period of no change the 2PS is more stable than the MSM reducing the MSE ($R^{(1)}$ measure) by approximately 4% while at the same time it responds faster to genuine growth changes. In addition it has proved more robust to errors in the nominated noise variance and can therefore be considered as a better forecasting model than the MSM.

Finally, the 33S attempts to improve the speed of growth response even further by droppingthe assumption made by the MSM namely that the probability of a process transition to state j (i.e.$\Pi^{(j)}$) at any time t is independent of the process state at time t-1. It therefore considers transition probabilities $\Pi^{(ij)}$ for a transition to state j at time t given that the process was in state i at time t-1. A new parameter, $\Pi^{(33)}$, is thus introduced which controls the speed of response produced by 33S to growth changes. Large values of $\Pi^{(33)}$ lead to faster growth response than the 2PS but run the risk of instability following a growth change. Such specification is therefore recommended only for applications (such as monitoring) where forecasting is not as important as fast detection of real changes. In forecasting applications the largest $\Pi^{(33)}$ value which can be tolerated is approximately .4 but this leads to a system whose performance is better than the MSM but worse than the 2PS.

CHAPTER 9

CONCLUSIONS

In the early 70's Harrison and Stevens (H/S [17, 18, 19]) made a major contribution to the area of statistical forecasting. They developed a Bayesian approach pioneering the use of Kalman filters in time series forecasting. They showed how previously accepted techniques can be restructured using a Markov formulation of time series. This structural change covers all existing linear time series models and with the introduction of likelihood methods the Bayesian approach provides a means of learning and communication. The advantages offered by this approach are numerous and an attempt has been made in chapter 1 to summarise them.

The prime objective of this thesis has been to investigate and examine methods of overcoming a number of theoretical and practical difficulties which arise in the application of Bayesian forecasting models and in particular of the Multi State Model (MSM). This model possesses all the advantages offered by the Bayesian approach and in addition has the ability to recognise quickly and respond appropriately to typical types of discontinuities in the process parameters. Chapter 3 attempted a full description of the mechanics of this model.

Chapter 4 examined the problems of assessing the performance produced by the MSM and showed that the traditional MSE criterion can be misleading when used for a process exhibiting a number of discontinuities. In addition the MSE is often meaningless since it does not allow any qualitative comparisons between different system responses. It was argued that no single criterion can possibly reflect the differences in response after points of discontinuities and therefore the performance of the MSM should be viewed as four dimensional with each dimension measuring the system's response to one of the four modelled process states. Four such measures $R^{(j)}$ for j = 1, 2, 3, 4 (corresponding to "no change", "outlier", "growth change" and "step change" states) were proposed two of which are quantitative and the other two qualitative. This it was argued, in contrast to any alternative set of purely quantitative criteria, makes comparisons of the system's behaviour more meaningful and facilitates subjective interpretation and judgement of the different responses.

The behaviour of the MSM can be controlled by the choice of a set of 7 system parameters (SSP = $\{\lambda_2 , \lambda_3 , \lambda_4 , \Pi^{(1)},\Pi^{(2)},\Pi^{(3)},\Pi^{(4)}\}$) Different SSP's lead to different responses and in chapter 5 an attempt was made to determine the sensitivity and robustness of the MSM in relation to the choice of an SSP. The set of responses produced by a particular SSP choice (the *standard* SSP) suggested by H/S has been used as a benchmark against which the responses of other SSP's have been compared. It was shown that the MSM is not robust to changes in the Π's from their *standard* values since this can lead to large instability during quiet periods. However it was also shown that equivalences between the Π's and the λ's exist in controlling the system to produce a particular response and it was therefore suggested that the best way of interacting with the system is by varying the λ's only with the Π's fixed at their *standard* values. The effect of the choice of λ's on the performance measures $R^{(j)}$ and its robustness to different types and sizes of discontinuities was then examined in section 5.3. Comparison and presentation of the results proved to be a difficult task since for every *input* $\underline{\lambda}$ set ($\underline{\lambda} = \{\lambda_2, \lambda_3, \lambda_4\}$) there were 10 associated measures of performance. It was shown that the *standard* $\underline{\lambda}$ set ($\underline{\lambda}_s$) is a robust and well balanced set of parameters, small variations from which resulted in negligible change in performance while larger variations could lead to a worsening of performance across the range of discontinuities examined. Hence if the process one is interested in, is expected to exhibit all the types of discontinuities considered (i.e. outliers, growth changes and step changes) no significant improvement in performance is possible to that produced by $\underline{\lambda}_s$. However in some special cases where one or more discontinuities are not expected to occur in the process under observation, significant improvements in performance can be made by choosing the λ's such as to remove the system's ability to recognise the irrelevant discontinuities.

Implicit in the analysis so far was a knowledge of the process noise variance V_ε. In practice however this is unknown and an estimate $V_{\varepsilon,N}$ must be nominated initially and $V_{\varepsilon,N}$ is therefore expected to be in some degree of error. Section 5.4 aimed at testing the robustness of the MSM responses to departures of $V_{\varepsilon,N}$ from V_ε. The results obtained suggest that underestimation of V_ε leads to substantial instability and spoils the robustness of the system to $\underline{\lambda}_s$. Furthermore, overestimation of V_ε leads to serious bias in the estimates of the process level (μ_t) following a medium sized step change and generally reduces the speed of response to changes. In view of the above it follows that V_ε is critical to the system's performance and therefore methods for its on line estimation are vital for successful implementation of the MSM in real data.

Two such procedures, FRM (fixed range method) and VRM (variable range method) were proposed in chapter 6 and tested on both real and artificial data. It was shown that the effect of the initial estimate of V_ε on the on line estimate at time t, $\hat{V}_{\varepsilon,t}$, dies away very quickly and $\hat{V}_{\varepsilon,t}$ is then wholly determined by the actual process observations, fluctuating reasonably close to the true value. The methods proved to be robust to all the different types and sizes of discontinuities typically found in real data and therefore both FRM and VRM overcome the problems (discussed in chapter 5) suffered by the MSM when it is operating with a fixed $V_{\varepsilon,N}$ which is either underestimating or overestimating the true value of the noice variance. In addition FRM and VRM respond satisfactorily to discontinuities in the noise variance which often occur in practice from circumstances beyond our control or expectations. However, the VRM has a significant computational advantage over FRM, requiring less than $\frac{1}{3}$ of the CPU time taken by the latter method and should therefore be preferred.

In chapter 7 the problem of on line variance estimation in the steady state model (SSM) was examined and it was shown that single state Bayesian models such as the SSM can produce misleading distributional information (which are very important for decision making) if fixed estimates of the process variances are used in the Kalman filter. The problem arises from the fact that the Kalman filter assumes an exact knowledge of these variances which must be known outside the system. It was therefore argued that on line variance estimation procedures are essential in order for the SSM to produce meaningful and near optimal distributional information. It was shown that for steady state processes where the variance ratio r_μ ($= V_\varepsilon / V_\mu$) is of the order of 10 or greater (implying an optimal α for an EWMA model of .27 or smaller) a good approximation to the optimal distributional information could be achieved by a simple modification of FRM and VRM which were originally developed for the MSM. In particular, the VRM is a very efficient method of estimating variance on line for single state Bayesian models (of any generality) or indeed with minor modifications in place of the usual methods based on average squared error or mean absolute deviation found in many stock control (EWMA type) systems.

However for $r_\mu < 10$ (EWMA $\alpha > .27$) better forecasting performance can be achieved by joint on line estimation of both V_ε and r_μ. Hence two more methods were proposed (CIM and CVM). CIM was tested on both artificial data and a read data series (B/J Data) reported by Box and Jenkins [3]. The final on line posterior estimates for B/J Data produced by CIM were very similar to those estimated off line by B/J. CVM in contrast to all previous methods is based on the assumption that V_ε and r_μ are absolutely constant through time. CVM exploits the limiting properties of the SSM and assumes a *discrete* prior probability distribution reflecting our uncertainty about r_μ. It was shown to produce on line maximum likelihood estimates of V_ε and r_μ. Its performance was illustrated using artifical data and it was also compared with the performance produced by a *continuous* method recently proposed by Leonard and Harrison (L/H [30]). The comparison showed that CVM produces practically identical forecasting performance while being computationally much more efficient.

Finally in chapter 8 it was aimed to test alternative formulations of the MSM in an attempt to improve both computational efficiency and its performance especially to abrupt growth changes. It was pointed out that expansion of matrix operations leads to significant cost saving by reducing the CPU time by a factor of the order of 30. Three alternative formulations to the MSM were then proposed (1PS, 2PS, 33S) and compared to each other. The 1PS behaves in a very similar manner to the MSM but reduces further the computational requirements by one half. Similar cost savings are achieved by the 2PS. However, the latter produces a significantly better set of responses than the MSM. It was shown that the 2PS improves the stability of the system during quiet periods as measured by the MSE ($R^{(1)}$ measure) which is reduced by approximately 4% while at the same time producing a faster response to growth changes. It was also shown to be more robust than the MSM to departures in $V_{\varepsilon,N}$ from the true noise variance V_{ε}. The 33S proved to be an improvement over the MSM but inferior to 2PS as far as forecasting applications are concerned. It was suggested however that 33S might have a use in a monitoring role where fast detections of real changes is more important than the stability of the system just after the change.

Further research might be worthwhile along the following three areas.

(i) The 2PS proved to have the best responses on the type of data examined but needs to be demonstrated on many real series. Furthermore it may be possible to extend it into a *three-point system* and define additional states modelling autocorrelations or other expected relationships.

(ii) Comparisons of different models and approaches such as Bayesian models, Box-Jenkins models, Adaptive filtering, exponential smoothing etc. have been made (one example being Fildes [13]) based on mean squared error. The disadvantages and risk involved in such comparisons have been discussed previously and it may be interesting to compare different systems using a set of performance measures like $R^{(j)}$ proposed in chapter 4.

(iii) The Kalman filter recurrence relationships assume the noise variance V_{ε} to be known. Whilst measures for tracking it have been developed in this work, explicit models for the movement of V_ε might be interesting. It may be possible for example to represent the observation noise ε_t as having come from a process whose parameters undergo all the types of movement and discontinuities associated in this work with the process parameters μ_t and β_t. Consequently a new range of *multi state models* for the noise variance could be developed.

APPENDIX A

NOTATION - ABBREVIATIONS - OPERATORS

This appendix contains a list of English and Greek symbols and abbreviations in alphabetical order. These are given in sections A.1, A.2 and A.3 respectively. Section A.4 lists some standard operators which have been used in the thesis without formal definition.

A.1 English symbols

Symbol	Section in which the symbol is first used or defined	
A_1	2.4	used to denote the smoothing constants of
A_2	"	conventional linear growth models e.g.Holt's
b_{t-1}	"	best estimate of the growth given D_{t-1}
$\underline{C}_t$	1.4	var/covariance matrix for $\underline{m}_t$
$c_{11,t-1}$	2.4	variance on the estimate of the level (m_{t-1})
$c_{22,t-1}$	"	" " " " " " growth (b_{t-1})
$c_{12,t-1}$	"	covariance of m_{t-1} and b_{t-1}
D_t	1.4	set of all data up to and including time t
e_t	2.3	one step ahead forecast error at time t
$\underline{F}_t$	1.4	vector defining how the process is observed
$\underline{G}$	1.4	matrix representing deterministic changes in $\underline{\Theta}_t$
k_2	4.2.2	constant determining the size of an "outlier"
k_3	4.2.3	" " " " " a "growth change"
k_4	4.2.4	" " " " " a "step change"

$\underline{m}_t$	1.4	best estimate of $\underline{\theta}_t$ at time t
m_{t-1}	2.4	best estimate of the level given D_{t-1}
$M_t^{(j)}$	3.2	model which at time t characterises (models) state j by a particular combination of $V_\varepsilon^{(j)}$, $V_\mu^{(j)}$, $V_\beta^{(j)}$ for j = 1, 2, 3, 4.
p_L	6.2.5	lower limit on probabilities
r_μ	2.3	constant true ratio (V_ε / V_μ)
r_β	2.4	" " " (V_μ / V_β)
$r_{\mu,N}$	2.3	initially fixed nominated estimate of r_μ
$r_{\beta,N}$	2.4	" " " " " r_β
$r_{\mu,t}$	7.3	dynamic ratio $(V_{\varepsilon,t}/V_{\mu,t})$
$\hat{r}_{\mu,t}$	7.3	best estimate of $r_{\mu,t}$ at time t
$R^{(j)}$	4.1	criterion measuring system performance to state j
V_ε	2.3	constant variance of ε_t for all t
V_μ	2.3	" " " $\delta\mu_t$ " " "
V_β	2.4	" " " $\delta\beta_t$ " " "
$V_{\varepsilon,N}$	2.3 & 3.4.3	initially fixed nominated estimate of V_ε
$V_{\mu,N}$	2.3	" " " " " V_μ
$V_{\beta,N}$	2.4	" " " " " V_β
V_t	1.4	Variance of ε_t at time t
$V_\varepsilon^{(j)}$	3.2	noise variance used to model state j
$V_\mu^{(j)}$	3.2	level disturbance variance used to model state j
$V_\beta^{(j)}$	3.2	growth " " " " " " "

$V_{\varepsilon}^{(k)}$	6.1	set of values in a V range for k = 1, 2, ... K
$V_{\varepsilon,t}$	6.1	unobservable true parameter representing the variance of the normal distribution with zero mean from which the observation noise at time t, ε_t, was obtained.
$\hat{V}_{\varepsilon,t}$	6.1	best estimate of $V_{\varepsilon,t}$ at time t
w	2.4	discount factor in the EWR model
$\underline{w}_t$	1.4	stochastic disturbance (in $\underline{\Theta}_t$) vector
$\underline{W}_t$	1.4	E ($\underline{w}_t$ $\underline{w}_t'$)
y_t	1.1	process observation at time t.

A.2 Greek symbols

β_t	2.4	unobservable true process growth at time t
$\delta\mu_t$	2.3	level (μ_t) disturbance at time t.
$\delta\beta_t$	2.4	growth (β_t) disturbance at time t
ε_t	1.4	observation noise at time t
$\underline{\Theta}_t$	1.4	unobservable vector of process parameters at time t
λ_2	3.2.2	parameter used to model an "outlier"
λ_3	3.2.3	" " " " a "growth change"
λ_4	3.2.4	" " " " a "step change"
$\underline{\lambda}$	5.3	the set of {λ_2 λ_3 λ_4}
$\underline{\lambda}_s$	5.3	set of standard λ values : {101, 1, 100}
μ_t	2.3	unobservable true process level at time t.
$\Pi^{(j)}$	3.3	constant probability of a process transition to state j.
σ	3.2.2	square root of V_{ε}
σ_N		square root of $V_{\varepsilon,N}$

A.3 Abbreviations

3M Data	6.2.1	Total (monthly) U.K. sales of 3M company
1PS	8.3	1 - point system
2PS	8.4	2 - point system
33S	8.5	"three-three" system
B/J	1.3	Box and Jenkins
B/J Data	6.2.1	concentration readings of a chemical process reported by B/J
C6 Data	6.2.1	Chapter 6 Data
C8 Data	8.3.1	Chapter 8 Data
CIM	7.3	Class I method of on line V_ε and r_μ estimation
CVM	7.4	constant variance method of on line V_ε and r_μ estimation
DLM	1.4	Dynamic Linear Model
EWMA	2.3	Exponentially weighted moving average
EWR	2.4	Exponentially weighted regression
FRM	6.1	Fixed range method
H/S	1.3	Harrison and Stevens
ICS	4.3.2	set of $\{SSP, V_{\varepsilon,N}\}$
L/H	7.4	Leonard and Harrison
LGM	2.4	linear growth model
MSE	1.5 & 4.4	Mean Square Error used mainly as a measure of stability during quiet periods.
MSM	1.4	Multi state model
RMSE	5.4.1	square root of average SSE
SSE	5.4.1	sum of squared errors

SSM	2.3	steady state model
SSP	3.4.4	set of system parameters $\{\underline{\lambda}, \Pi^{(j)}\}$ for $j = 1, 2, 3, 4$
V range	6.1	set of $V_{\varepsilon}^{(k)}$ values covering the likely value of $V_{\varepsilon,t}$
VRM	6.3	Variable range method

A.4 Operators

$E(x)$	:	expected value of x
$Var(x)$	:	variance of x
$Cov(x,y)$	:	covariance of x and y
$\underline{x}'$	:	transpose of $\underline{x}$
$(\underline{x})^{-1}$	:	inverse of $\underline{x}$
$p(x\|y)$	:	conditional probability of x given y
$L(x\|y)$	:	conditional likelihood of x given y

APPENDIX B

THE DERIVATION OF THE KALMAN FILTER EQUATIONS

Consider the general DLM formulation given by (2.2.1) and (2.2.2):

$$y_t = \underline{F}_t \, \underline{\Theta}_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, V_t) \tag{B.1}$$

$$\underline{\Theta}_t = \underline{G} \, \underline{\Theta}_{t-1} + \underline{w}_t \qquad \underline{w}_t \sim N(\underline{0}, \underline{W}_t) \tag{B.2}$$

Assuming that prior to y_t our information about the parameter vector is of the form,

$$(\underline{\Theta}_{t-1} \mid D_{t-1}) \sim N(\underline{m}_{t-1} \, , \, \underline{C}_{t-1}) \tag{B.3}$$

our aim is to show that as soon as y_t and $\underline{F}_t$ become known then (B.1), (B.2) and (B.3) imply that the posterior distribution is also normal,

$$(\underline{\Theta}_t \mid D_t) \sim N(\underline{m}_t \, , \, \underline{C}_t)$$

where $\underline{m}_t$, $\underline{C}_t$ are as given by (2.2.6) and (2.2.7) respectively, in chapter 2.

From (B.1), (B.2), and (B.3) it is easy to derive the following using the additive property of means and Normal variances:

$$E(\underline{\Theta}_t) = \underline{G}\ \underline{m}_{t-1} \qquad (B.4)$$

$$E(y_t) = \underline{F}_t\ \underline{G}\ \underline{m}_{t-1} \qquad (B.5)$$

$$\text{Var}\ (\underline{\Theta}_t) = \underline{G}\ \underline{C}_{t-1}\ \underline{G}' + \underline{W}_t \qquad (B.6)$$

$$\text{Var}\ (y_t) = \underline{F}_t \text{Var}(\underline{\Theta}_t) \underline{F}_t' + V_t \qquad (B.7)$$

$$\left.\begin{array}{ll} \text{Let} & \hat{y}_t = E(y_t) \\ & \underline{R}_t = \text{Var}\ (\underline{\Theta}_t) \\ \text{and} & \hat{Y}_t = \text{Var}\ (y_t) \end{array}\right\} \qquad (B.8)$$

then $\hat{y}_t$, $\underline{R}_t$ and $\hat{Y}_t$ are given by (B.5), (B.6) and (B.7) respectively.

Consider now the covariance between $\underline{\Theta}_t$ and y_t ,

$$\text{Covar}\ (\underline{\Theta}_t\ , y_t) = E\{[\underline{\Theta}_t - E(\underline{\Theta}_t)]\ [y_t - E(y_t)]'\} \qquad (B.9)$$

but since $y_t = \underline{F}_t\ \underline{\Theta}_t + \varepsilon_t$ from (B.1), it follows that

$E(y_t) = \underline{F}_t\ E(\underline{\Theta}_t)$ and therefore $[y_t - E(y_t)]' = [\underline{F}_t\ \underline{\Theta}_t + \varepsilon_t - \underline{F}_t\ E(\underline{\Theta}_t)]'$

$$= [\underline{\Theta}_t - E(\underline{\Theta}_t)]'\ \underline{F}_t' + \varepsilon_t$$

Substituting this result in (B.9) and using the fact that $\underline{\Theta}_t$ is independent of the purely random noise term ε_t, it follows that,

$$\text{Covar} \quad (\underline{\Theta}_t, y_t) = \text{Var} \ (\underline{\Theta}_t) \ \underline{F}_t' = \underline{R}_t \ \underline{F}_t' \qquad \text{(B.10)}$$

since from (B.8) $\underline{R}_t = \text{Var} \ (\underline{\Theta}_t)$

We can now write the joint distribution of $\underline{\Theta}_t$ and y_t as follows using the results already obtained:

$$\left(\begin{bmatrix} \Theta_t \\ \\ y_t \end{bmatrix} \Bigg| D_{t-1}, \underline{F}_t \right) \sim N \left(\begin{bmatrix} \underline{G} \ \underline{m}_{t-1} \\ \\ \hat{y}_t \end{bmatrix} ; \begin{bmatrix} \underline{R}_t & \underline{R}_t \ \underline{F}_t' \\ \\ \underline{F}_t \ \underline{R}_t' & \hat{Y}_t \end{bmatrix} \right)$$

Standard statistical theory (e.g. Anderson /1/) gives the distribution of $\underline{\Theta}_t$ conditional on y_t as:

$$(\underline{\Theta}_t \mid y_t \ , D_{t-1}, \underline{F}_t) \sim N(\underline{m}_t \ , \underline{C}_t)$$

or $$(\underline{\Theta}_t \mid D_t) \sim N(\underline{m}_t \ , \underline{C}_t)$$

(since by definition $D_t = \{D_{t-1} \ , y_t, \underline{F}_t\}$)

where

$$\underline{m}_t = \underline{G}\,\underline{m}_{t-1} + \underline{R}_t\,\underline{F}'_t\;\hat{Y}_t^{-1}\;(y_t - \hat{y}_t)$$

$$\underline{C}_t = \underline{R}_t - \underline{R}_t\,\underline{F}'_t\,\hat{Y}_t^{-1}\,\underline{F}_t\,\underline{R}'_t$$

and using the notation for e_t , $\underline{A}_t$ given in (2.2.5) these simplify to:

$$\underline{m}_t = \underline{G}\,\underline{m}_{t-1} + \underline{A}_t\,e_t$$

$$\underline{C}_t = \underline{R}_t - \underline{A}_t\,\hat{Y}_t\,\underline{A}'_t$$

APPENDIX C

THE INCORPORATION OF SEASONAL AND PROMOTIONAL EFFECTS

C.1 A seasonal multi-state model

The forecasting system which will be discussed here represents an actual application of a Bayesian multi-state model to the total sales data of a London based multinational company specialising in surface coatings.

Their data exhibits several short term growth changes and it was felt that a Bayesian model might be appropriate producing a fast response to abrupt changes in growth while at the same time producing a stable response during "quiet" periods. This is the kind of trade off that makes traditional forecasting methods unsuitable for many real life applications.

The system developed will only be briefly outlined here but a full description including a computer program listing and detailed documentation is given by Cantarelis [9]. The model used is multiplicative in line with previous studies (e.g. Harrison and Scott [21]) which have found it to be more suitable than an additive one for most seasonal data.

It incorporates twelve seasonal factors superimposed on a linear growth process and assumes the following form for the actual observations:

$$\left.\begin{aligned} y^*_t &= \mu^*_t \cdot \rho^*_t \cdot \varepsilon^*_t \\ \mu^*_t &= \mu^*_{t-1} \cdot \beta^*_t \cdot \delta\mu^*_t \\ \beta^*_t &= \beta^*_{t-1} \cdot \delta\beta^*_t \\ \rho^*_{t+i} &= \rho^*_{t+i-1} \cdot \delta\rho^*_{t+i} \quad \text{for } i = 0, 1, 2, \ldots. 11 \end{aligned}\right\} \tag{C.1.1}$$

where

y^*_t = actual observation at time t

μ^*_t = level of the process at time t

β^*_t = growth of the process at time t

ρ^*_{t+i} = seasonal factor of that month which is i months away from the current time = t month. So ρ^*_t (i.e. i = 0) is in fact the seasonal factor of the current month, ρ^*_{t+1} is the seasonal factor of the next month etc.

ε_t^* = log-normal observation noise

$\delta\mu_t^*$ = log-normal level disturbance

$\delta\beta_t^*$ = log-normal growth disturbance

$\delta\rho_{t+i}^*$ = log-normal seasonal disturbance

By taking logarithms and using the obvious notation,

$y_t = \log y_t^*$

$\mu_t = \log \mu_t^*$

e.t.c.

the model is easily converted into DLM form:

$$\left.\begin{array}{lll} \text{Observation eq}^{\underline{n}} & : & y_t = \mu_t + \rho_t + \varepsilon_t \\ \text{System eq}^{\underline{n}} & : & \mu_t = \mu_{t-1} + \beta_t + \delta\mu_t \\ & & \beta_t = \beta_{t-1} + \delta\beta_t \\ & & \rho_{t+i} = \rho_{t+i-1} + \delta\rho_{t+i} \\ & & \text{for } i = 0, 1, 2, \ldots . \; 11 \end{array}\right\} \quad \text{(C.1.2)}$$

with $\varepsilon_t \sim N(0, V_\varepsilon)$

$\delta\mu_t \sim N(0, V_\mu)$

$\delta\beta_t \sim N(0, V_\beta)$

$\delta\rho_{t+i} \sim N(0, V_\rho)$ for all i, and subject to the constraint:

$$\sum_{i=0}^{11} \rho_{t+i} = 0 \qquad \text{for all } t \qquad \text{(C.1.3)}$$

which ensures that the sum of the log-seasonals is zero over the complete cycle of 12 months. In other words the constraint implies that the seasonal peaks must cancel out the troughs. Clearly for (C.1.3) to hold for all t the sum of disturbances $\delta\rho_{t+i}$ must also sum to zero over the 12 month cycle, i.e.

$$\sum_{i=0}^{11} \delta\rho_{t+i} = 0 \qquad \text{for all } t \qquad \text{(C.1.4)}$$

The relation to the general DLM formulation given in section (1.4) is as follows:

Observation eq$^{\underline{n}}$: $y_t = \underline{F}_t \, \underline{\Theta}_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, V_t)$

System eq$^{\underline{n}}$: $\underline{\Theta}_t = \underline{G} \, \underline{\Theta}_{t-1} + \underline{w}_t \qquad \underline{w}_t \sim N(0, \underline{W}_t)$

with

$$n = 14$$

$$\underline{F}_t = (1, 0 \mid 1, 0, 0, \ldots. 0)$$

$$\underline{\Theta}_t = (\mu_t, \beta_t \mid \rho_t, \rho_{t+1}, \rho_{t+2}, \ldots.. \rho_{t+11})'$$

$$\underline{\Theta}_{t-1} = (\mu_{t-1}, \beta_{t-1} \mid \rho_{t+11}, \rho_t, \rho_{t+1}, \ldots \rho_{t+10})'$$

$$\underline{G} = \left[\begin{array}{cc|c} 1 & 1 & \\ & & \underline{0} \\ 0 & 1 & \\ \hline \underline{0} & & \underline{G}^* \end{array}\right]$$

$$\underline{G}^* = \left[\begin{array}{ccccccccc} 0 & 1 & 0 & & & & & & \\ & 0 & 1 & 0 & & & & & \\ & & 0 & 1 & 0 & & & & \\ & & & & & & & & \\ & & & & & & 0 & 1 & 0 \\ & & & & & & 0 & 0 & 1 \\ 1 & 0 & 0 & & & & 0 & 0 & 0 \end{array}\right]$$

and $\underline{V}_t = V_\varepsilon$

Before we look at $\underline{W}_t$ which must be of a particular form we look at the information required initially for the parameter vector $\underline{\theta}_o$:

$$(\underline{\theta}_o \mid D_o) \sim N(m_o , \underline{C}_o)$$

where

$$\underline{m}_o = \left[\begin{array}{c} m_t \\ b_t \\ \hdashline S_{t+11} \\ S_t \\ S_{t+1} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ S_{t+10} \end{array}\right] = \left[\begin{array}{cccc} \text{our} & \text{best} & \mu_o & \text{estimate} \\ " & " & \beta_o & " \\ \hdashline " & " & \rho_{11} & " \\ " & " & \rho_o & " \\ " & " & \rho_1 & " \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ " & " & \rho_{10} & " \end{array}\right]$$

$$\text{and } \underline{C}_o = \left[\begin{array}{cc|c} c_{11,0} & 0 & \\ & & \underline{0} \\ 0 & c_{22,0} & \\ \hline & \underline{0} & \underline{C}^*_o \end{array}\right]$$

$$\text{with } \underline{C}^*_o = \left[\begin{array}{cccc} c_{33,0} & & & -\frac{c_{33,0}}{11} \\ & c_{33,0} & & \\ & & \ddots & \\ & \frac{-c_{33,0}}{11} & & c_{33,0} \end{array}\right]$$

where $c_{11,0} = \mathrm{Var}\ (\mu_o \mid D_o)$

$c_{22,0} = \mathrm{Var}\ (\beta_o \mid D_o)$

$c_{33,0} = \mathrm{Var}\ (\rho_i \mid D_o)$ for all i .

Initially the covariance between μ_o and β_o is taken equal to zero but negative covariances for the seasonals are required in order to make the rows and columns of $\underline{C}^*_o$ sum to zero. This negative correlation is the result of the restriction (C.1.3) which implies that ρ_{t+i} are not independent since their sum must be zero.

More complicated forms of C^*_o could also be specified subject to the condition that its rows and columns sum to zero. This is necessary in order that $\underline{A}_t$ in the Kalman filter will be such as to ensure that our estimates of the seasonal factors also sum to zero, i.e.

$$\sum_{i=o}^{11} S_{t+i} = 0 \quad \text{for all } t$$

Also from the Kalman filter, it follows that for $\underline{C}^*_1$ (updated version of $\underline{C}^*_o$) to have rows and columns summing to zero it is necessary for the rows and columns of $\underline{W}^*$ to sum to zero also, where

$$
\underline{W}_t = \left[\begin{array}{cc|c} \mathrm{Var}\,(\delta\mu_t) + \mathrm{Var}\,(\delta\beta_t) & \mathrm{Var}\,(\delta\beta_t) & \\ & & \underline{0} \\ \mathrm{Var}\,(\delta\beta_t) & \mathrm{Var}\,(\delta\beta_t) & \\ \hline \multicolumn{2}{c|}{\underline{0}} & \underline{W}^* \end{array}\right]
$$

and the form of $\underline{W}^*$ used is as follows:

$$
\underline{W}^* = \left[\begin{array}{c|c} 1 & \cdots\cdots \ -\frac{1}{11} \ \cdots\cdots \\ \hline \vdots & \\ -\frac{1}{11} & \frac{1}{121} \\ \vdots & \end{array}\right] \cdot V_\rho
$$

Hence the structure of the $\underline{C}_o$ and $\underline{W}_t$ matrices is such that the seasonal estimates of ρ_{t+i} sum to zero at all times.

The complete model assumes that at any one time the series is in one of the four states described in Table 1.1 of section 1.4. That is, a model $M^{(j)}$ characterises state j at time t through distinct combinations of the noise and disturbance variances, i.e.

$$M^{(j)} = \{\underline{F}_t , \underline{G} , V_t^{(j)} , \underline{W}_t^{(j)}\} \quad j = 1, 2, 3, 4.$$

It is further assumed that the system enters these states at random independent of time and process history according to an initially specified probability set $\Pi^{(j)}$ where $\Pi^{(j)}$ denotes the probability of a process transition to state j at time t. The set of all these noise, disturbance and probability parameters $\{V_t^{(j)} , \underline{W}_t^{(j)} , \Pi^{(j)}\}$ represent beliefs about the environment that controls the data series and specify the nature and magnitude of discontinuities that may be expected to take place. The exact way of specifying these parameters in order to model the different states as well as the procedure for revising information and producing forecasts is described in Chapter 3 and in [9].

C.2 A Promotional Model

The model outlined here has been produced by Wright [45] for a Liverpool based company manufacturing savouries. The problem was that marketing activities such as advertising promotions when first launched have a major effect and introduce a great deal of uncertainty. However some prior knowledge of their timing and likely magnitude is available and could be incorporated by a Bayesian model. It was felt that advertising and consumer offers would have the sort of effect on the demand, described by figure C.2.1 below:

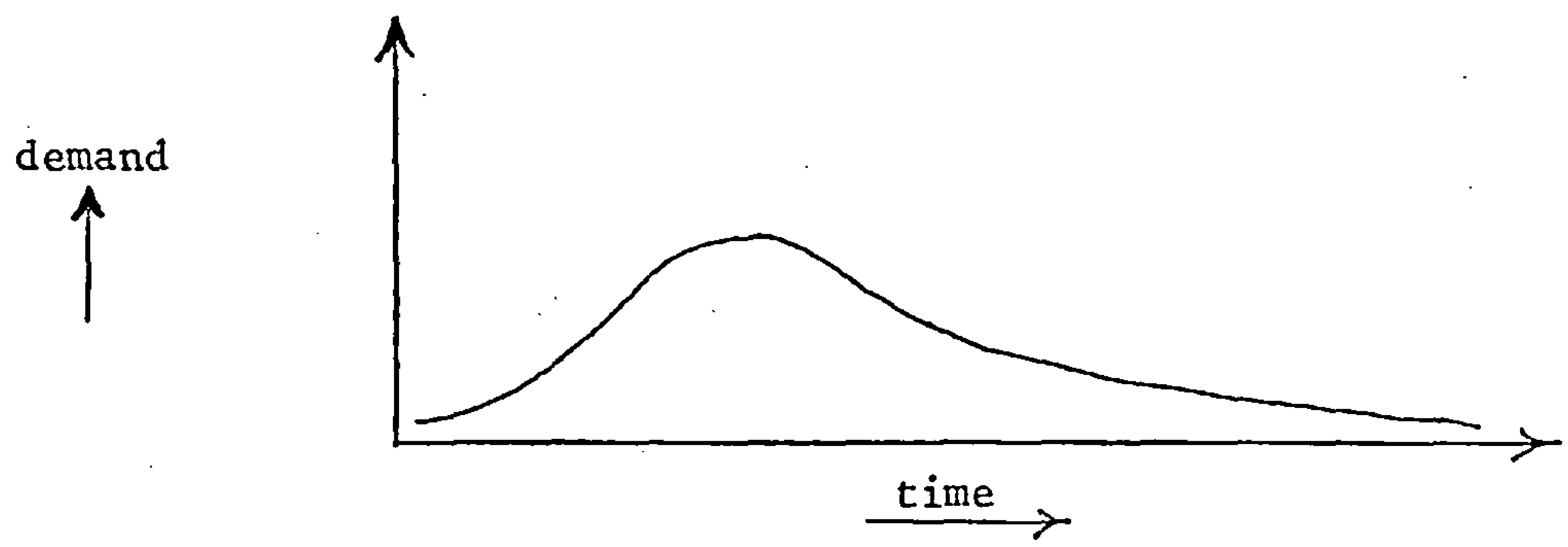


Figure C.2.1

That is, a straightforward increase in demand over a certain period, where the length and magnitude of this increase is dependent on factors such as the size of promotion, the type pf promotion and the market sector at which this promotion is aimed.

A different marketing approach also considered was price rises. A price rise was thought to result in an increase in demand for the period prior to the price rise, as retailers stock up at the old price, followed by a drop in demand in the next time period, partly because the retailers are already stocked up and partly because of the increase in the price of the product. This effect is shown in Figure C.2.2. below:

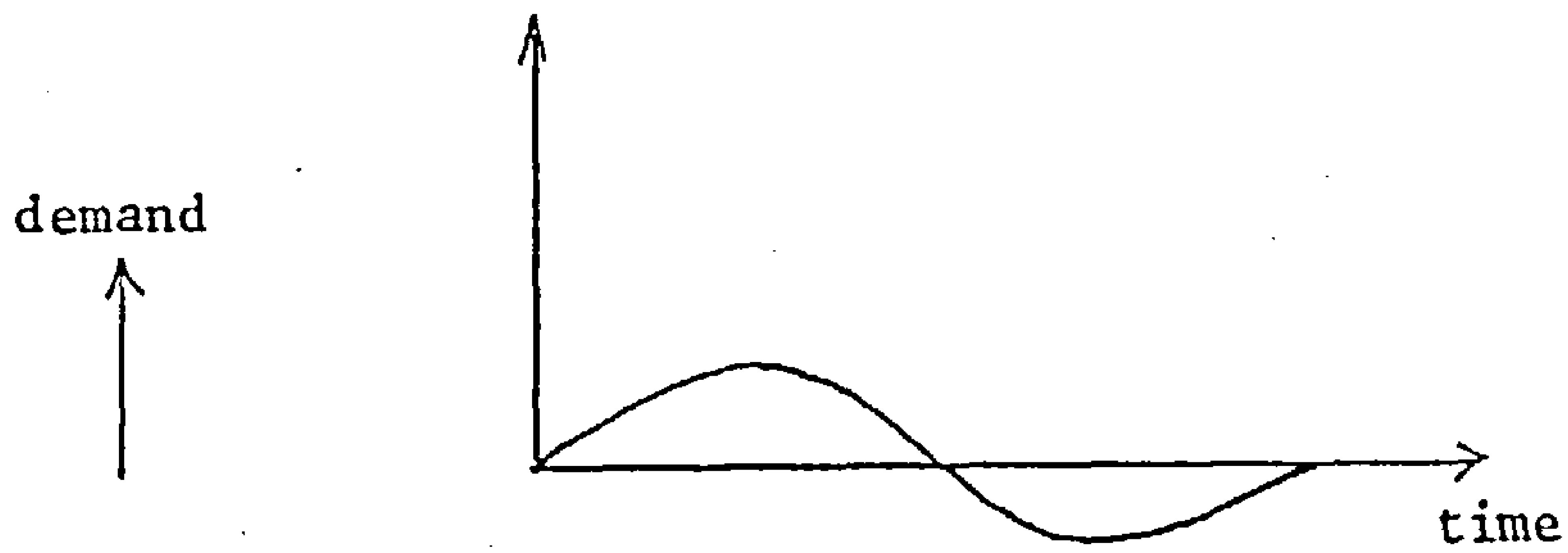


Figure C.2.2.

The final model produced incorporated these two basic effects of (i) consumer promotions and (ii) price rises, superimposed on a linear growth seasonal model similar to that described in C.1 except that only one state was modelled rather than four. The range of the estimated magnitudes for the promotional effects that were used was 2% to 5% and typical estimates would be:

(i) effect of a 4 period (one period = 4 weeks) promotion:

+ 2% + 5% + 5% + 3%

(ii) two period effect of a price rise:

+ 3% - 5%

where + 5% means that the demand is expected to be 5% higher than "normal" demand which would have been realised if no promotion activity had taken place.

The observation and system equations now take the following form:

$$\left.\begin{aligned} y_t &= \mu_t + \rho_t + \chi_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \beta_t + \delta\mu_t \\ \beta_t &= \beta_{t-1} + \delta\beta_t \\ \rho_{t+i} &= \rho_{t+i-1} + \delta\rho_{t+i} \\ \chi_{t+i} &= \chi_{t+i-1} + \delta\chi_{t+i} \end{aligned}\right\} \quad (C.2.1)$$

for i = 0, 1, 2, ... 12 since the model views a year as 13 4-week periods. The model is similar to that of (C.1.2) except that 13 parameters (χ_{t+i} $i = 0, 1, 2 \ldots 12$) have been added to take into account the promotional effects for up to 13 periods ahead. The subscript of χ_{t+i} has the same interpretation with that given in C.1 for the seasonal parameters ρ_{t+i} . The promotional factors are incorporated into an extended parameter vector $\underline{\theta}_t$ which at time t is estimated by $\underline{m}_t$. The structure of $\underline{m}_t$ is as follows:

$$\underline{m}_t = \left[\begin{array}{c} \left.\begin{array}{c} m_t \\ b_t \end{array}\right\} \\ \hline \left.\begin{array}{c} S_t \\ S_{t+1} \\ \vdots \\ S_{t+12} \end{array}\right\} \\ \hline \left.\begin{array}{c} q_t \\ q_{t+1} \\ \vdots \\ q_{t+12} \end{array}\right\} \\ \hline 0 \\ \hline U_1 \\ U_2 \\ U_3 \\ U_4 \end{array}\right]$$

estimates of μ_t and β_t

estimates of the true but unknown seasonal factors ρ_{t+i} $i = 0, 1, 2 \ldots .12$

estimates of the true but unknown promotional effects χ_{t+i} $i = 0, 1, 2 \ldots 12$

The final four positions of $\underset{-}{m}_t$ are there to accommodate the updated estimates of the last four promotional effects so that when the promotion has finished the initial estimates can be compared to these updated estimates which should help in future estimation of promotional effects. The reason for the zero in the fifth position from the end of $\underset{-}{m}_t$ will become clear when the structure of $\underset{-}{G}$ is given below. The $\underset{-}{F}_t$ matrix determines the parameters to be included in the observation equation and is therefore of the following form:

$$\underset{-}{F}_t = [\, 1 \quad 0 \;\vdots\; 1, \text{ twelve zeros } \;\vdots\; 1 \text{ , eighteen zeros} \,]$$

thus selecting the level (m_t), the current seasonal factor (S_t) and the current promotional effect (q_t). The matrix $\underset{-}{G}$ can be represented by two non zero matrices $\underset{-}{G}_A$ and $\underset{-}{G}_B$:

$$\underset{-}{G} = \left[\begin{array}{c|c} \underset{-}{G}_A & \underset{-}{0} \\ \hline \underset{-}{0} & \underset{-}{G}_B \end{array}\right]$$

where $\underset{-}{G}_A$ is identical to the $\underset{-}{G}$ matrix of the previous model in C.1 except that it is now a (15 x 15) matrix instead of (14 x 14) since it handles 13 seasonal factors (plus the level and growth) instead of 12 used by the previous model.

$\underset{-}{G}_B$ has the following special structure:

$$\underline{G}_B = \begin{bmatrix} 0 & 1 & 0 & & & & & & & & & \\ & 0 & 1 & 0 & & & & & & & & \\ & & 0 & 1 & 0 & & & & & & & \\ & & & & \ddots & & & & & & & \\ & & & & & 0 & 1 & 0 & & & & \\ & & & & & & 0 & 0 & 0 & & & \\ & & & & & & & 0 & 1 & 0 & & \\ & & & & & & & & 0 & 1 & 0 & \\ & & & & & & & & & 0 & 0 & 1 \\ 1 & 0 & 0 & & & & & & & 0 & 0 & 0 \end{bmatrix}$$

$\underline{G}_B$ is an (18 x 18) matrix with a structure similar to that required by the seasonal factors of the model described in C.1 in order to be rotated. The only difference is that the fifth row from the bottom contains only zeros and as a result this row removes the updated promotional estimate used four periods back. This is necessary in order to eventually remove all the promotional factors since these are not cyclical effects (like the seasonal factors) but transient. Finally the form of the $\underline{W}_t$ matrix is,

$$\underline{W}_t = \left[\begin{array}{c|c} \underline{W}_A & \\ \hline & \underline{W}_B \end{array}\right]$$

where $\underline{W}_A$ is identical to $\underline{W}_t$ of the previous model except that it provides for 13 seasonal factors rather than 12. $\underline{W}_B = [\ \underline{0}\ ]$ since it represents the long term uncertainty about χ_{t+i} and given that the maximum length of time that marketing promotions last is of the order of four periods (16 weeks) it is not very meaningful to make any statement about their long term uncertainty.

APPENDIX D

VARIANCE OF THE ONE STEP AHEAD FORECAST ERROR e_t IN THE SSM

At time t+1 the one step ahead forecast error is:

$$e_{t+1} = y_{t+1} - \hat{y}_{t+1} \qquad (D.1)$$

Using equations (2.3.1) and (2.3.2) this can be written as:

$$e_{t+1} = \mu_t + \delta\mu_{t+1} + \varepsilon_{t+1} - \hat{y}_{t+1}$$

Hence $E(e_{t+1}) = E(\mu_t) - E(\hat{y}_{t+1}) = m_t - \hat{y}_{t+1} = 0$

and $\mathrm{Var}\ (e_{t+1}) = V_\mu + V_\varepsilon + \mathrm{Var}\ (\hat{y}_{t+1})$ (D.2)

Using equations (2.3.6) and (2.3.7) we have:

$$\hat{y}_{t+1} = m_t = A_N \sum_{i=0}^{\infty} (1 - A_N)^i \, y_{t-i} \qquad (D.3)$$

From equations (2.3.1) and (2.3.2) y_{t-i} can be written as:

$$y_{t-i} = \mu_{t-i} + \varepsilon_{t-i} = \mu_t + \varepsilon_{t-i} - \sum_{j=0}^{i-1} \delta\mu_{t-j} \qquad (D.4)$$

If we now use β to denote $(1-A_N)$ we have from D.3 and D. 4

$$\hat{y}_{t+1} = (1-\beta) \sum_{i=o}^{\infty} \beta^{i}(\mu_t + \varepsilon_{t-i}) + (1-\beta) \sum_{i=o}^{\infty} \beta^{i} \left(\sum_{j=o}^{i-1} \delta\mu_{t-j}\right) \qquad \text{(D.5)}$$

Consider now the summation of the second term :

$$\sum_{i=o}^{\infty} \beta^{i} \sum_{j=o}^{i-1} \delta\mu_{t-j} = \sum_{i=o}^{\infty} \delta\mu_{t-i} \sum_{j=i+1}^{\infty} \beta^{j} = \sum_{i=o}^{\infty} \delta\mu_{t-i} \left(\frac{\beta^{i+1}}{1-\beta}\right)$$

Substituting this result in (D.5) gives:

$$\hat{y}_{t+1} = (1-\beta) \sum_{i=o}^{\infty} \beta^{i}(\mu_t + \varepsilon_{t-i}) + \sum_{i=o}^{\infty} \delta\mu_{t-i} \beta^{i+1}$$

Hence $\text{Var}\ (\hat{y}_{t+1}) = V_{\varepsilon}(1-\beta)^2 \sum_{i=o}^{\infty} \beta^{2i} + V_{\mu} \sum_{i=o}^{\infty} \beta^{2i+2} =$

$$= \frac{(1-\beta)^2\ V_{\varepsilon} + \beta^2\ V_{\mu}}{1-\beta^2}$$

Substituting this in (D.2) we finally get:

$$\text{Var}\ (e_{t+1}) = \frac{2V_{\varepsilon}(1-\beta) + V_{\mu}}{1-\beta^2} = \frac{2A_N\ V_{\varepsilon} + V_{\mu}}{A_N(2-A_N)}$$

APPENDIX E

MIXED DISTRIBUTIONS IN THE COLLAPSING PROCESS

In Section 3.3 it was seen that four bivariate normal distributions must be approximated by a single distribution. For simplicity we will restate the problem using slightly different notation and consider the case of one dimensional distributions. The multivariate results can then be derived in a similar way.

Let $f_i(\mu)$ for i = 1, 2, 3, 4 be four normal distributions with means m_i and variances C_i :

$$f_i(\mu) = (2\pi C_i)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\frac{(\mu-m_i)^2}{C_i}\right\} \qquad (E.1)$$

and let p_i i = 1, 2, 3, 4 be probabilities associated with $f_i(\mu)$ such that $\sum_{i=1}^{4} p_i = 1$. We now assume that the *mixed* distribution $f(\mu)$,

$$f(\mu) = \sum_{i=1}^{4} p_i f_i(\mu) \qquad (E.2)$$

can be approximated by a normal distribution $f^*(\mu)$ with mean m and variance C :

$$f^*(\mu) = (2\pi C)^{-\frac{1}{2}} \exp\left\{\frac{-(\mu-m)^2}{2C}\right\} \qquad (E.3)$$

where m is equal to the first moment of $f(\mu)$ about the origin and C is the second moment of $f(\mu)$ about its expected value.

The first moment of $f(\mu)$ taken about the origin can be written as:

$$E(f(\mu)) = E_1 = \sum_{i=1}^{4} p_i \int_{-\infty}^{+\infty} \mu\, f_i(\mu)\, d\mu$$

or

$$E_1 = \sum_{i=1}^{4} p_i \underset{f_i}{E} (\mu) = \sum_{i=1}^{4} p_i\, m_i$$

The second moment of $f(\mu)$ about its expected value E_1, can be written as follows:

$$E\left[f(\mu) - E_1\right]^2 = E_2 = \sum_{i=1}^{4} p_i \int_{-\infty}^{+\infty} (\mu - E_1)^2 f_i(\mu)\, d\mu$$

But

$$\int_{-\infty}^{+\infty} (\mu - E_1)^2 f_i(\mu)\, d\mu = \underset{f_i}{E} (\mu - E_1)^2 =$$

$$= \underset{f_i}{E} \left[(\mu - m_i) + (m_i - E_1)\right]^2 =$$

$$= \underset{f_i}{E} (\mu - m_i)^2 + 2(m_i - E_1) \underset{f_i}{E} (\mu - m_i) + (m_i - E_1)^2$$

$$= C_i + 0 + (m_i - E_1)^2$$

Hence $$E_2 = \sum_{i=1}^{4} p_i \{C_i + (m_i - E_1)^2\}$$

Finally our approximation for the mixed distribution is $f^*(\mu)$ as given by (E.3) with :

$$m = E_1 = \sum_{i=1}^{4} p_i m_i$$

and $$C = E_2 = \sum_{i=1}^{4} p_i \{C_i + (m_i - m)^2\}$$

The analogy between m, C and the collapsing process equations given by (3.3.12) is now obvious.

APPENDIX F

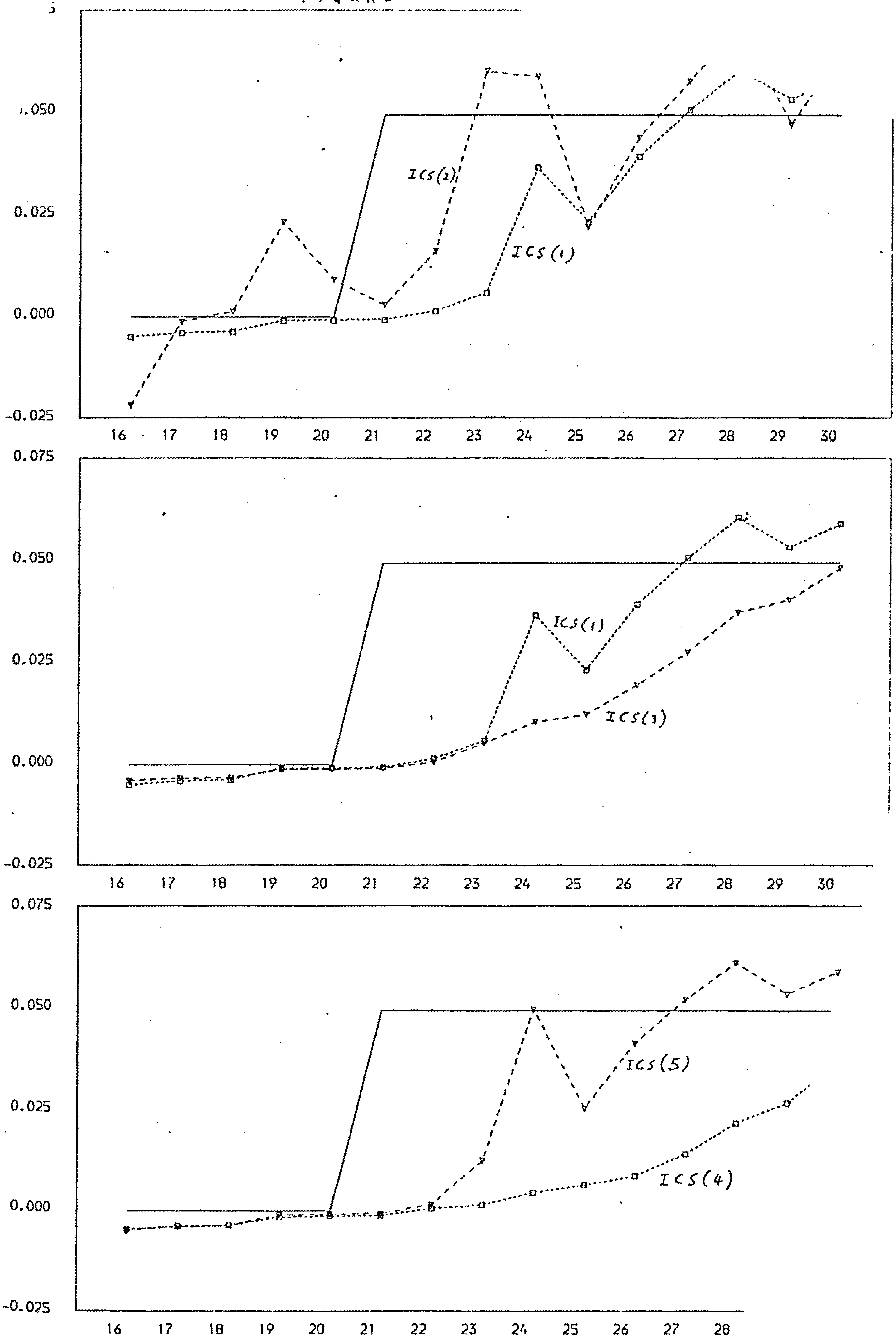
Generation of random Normal deviates

Random normal deviates $r_t \sim N(0, 1)$ have been generated using a NAG procedure (NAG stands for Nottingham Algorithm Group). This procedure generates r_t such that alternate elements of the sequence are determined from the following equations:

$$\left.\begin{aligned} r_{2t-1} &= \sin(2\pi x_{2t}) \cdot \left\{-2 \log_e (x_{2t-1})\right\}^{-\frac{1}{2}} \\ r_{2t} &= \cos(2\pi x_{2t}) \cdot \left\{-2\log_e (x_{2t-1})\right\}^{-\frac{1}{2}} \end{aligned}\right\}$$

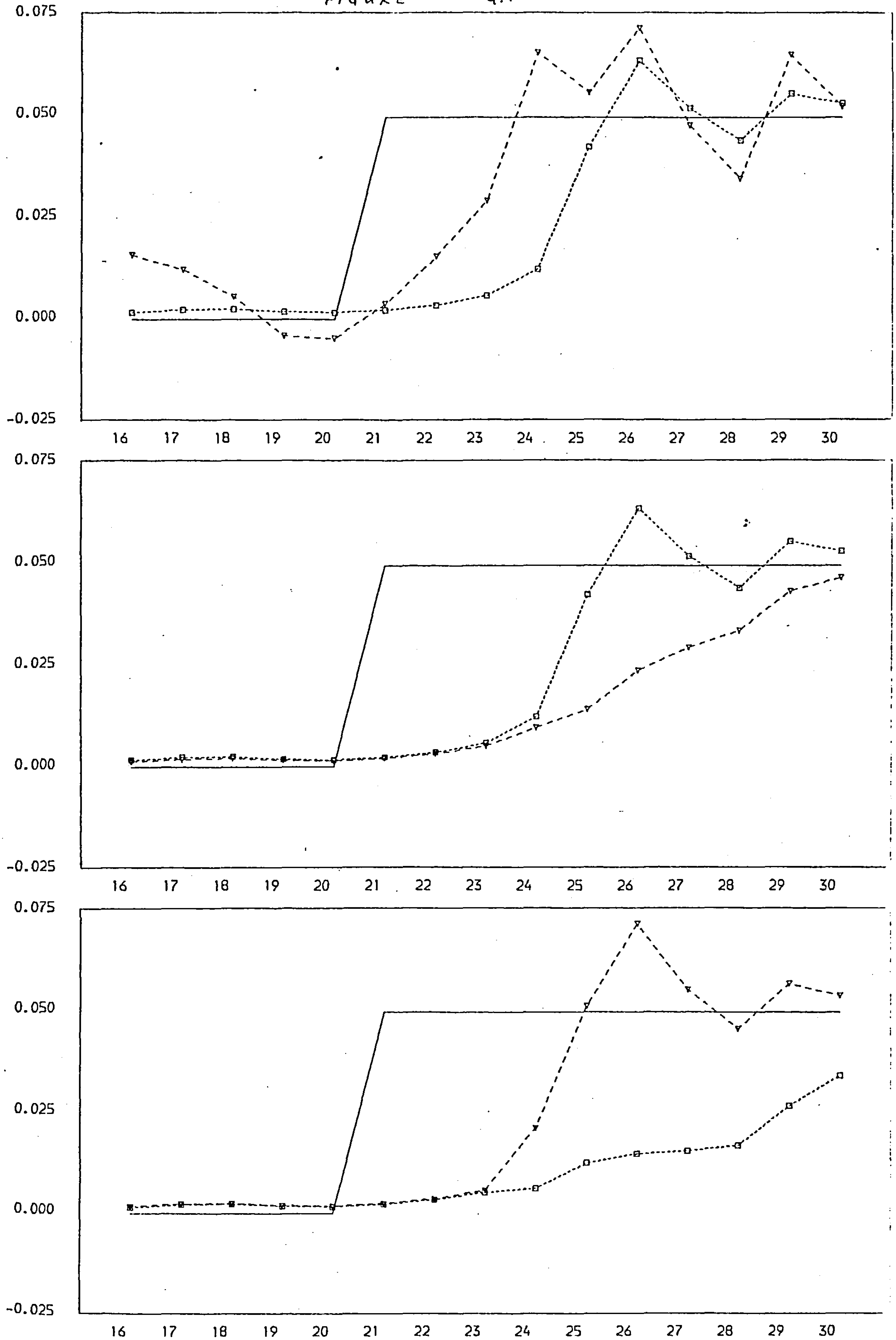
where x_t is a pseudo random number. For details of this random normal deviate generation routine, see Box and Muller [4].

Neave [49] pointed out that the sampling distribution obtained in this way can have local maxima while Golder and Settle [50] suggest a "two sequence" method whereby two sets of pseudo random uniform deviates $\{x_i\}$ and $\{x_i'\}$ produced by two multiplicative congruential generators are used in the Box-Muller formulation in place of x_{2t-1} and x_{2t}.



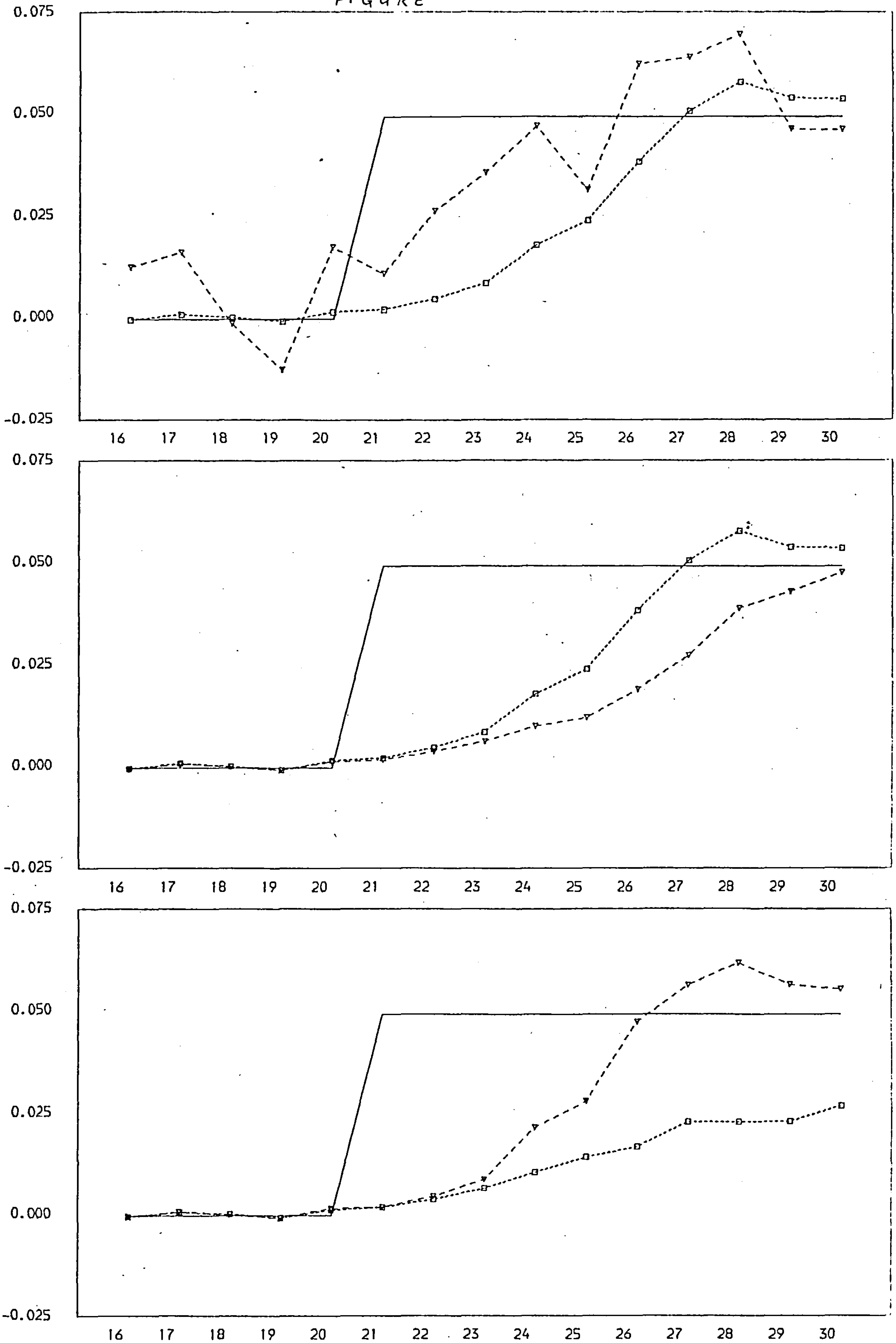
REALISATION 1

FIGURE G.1



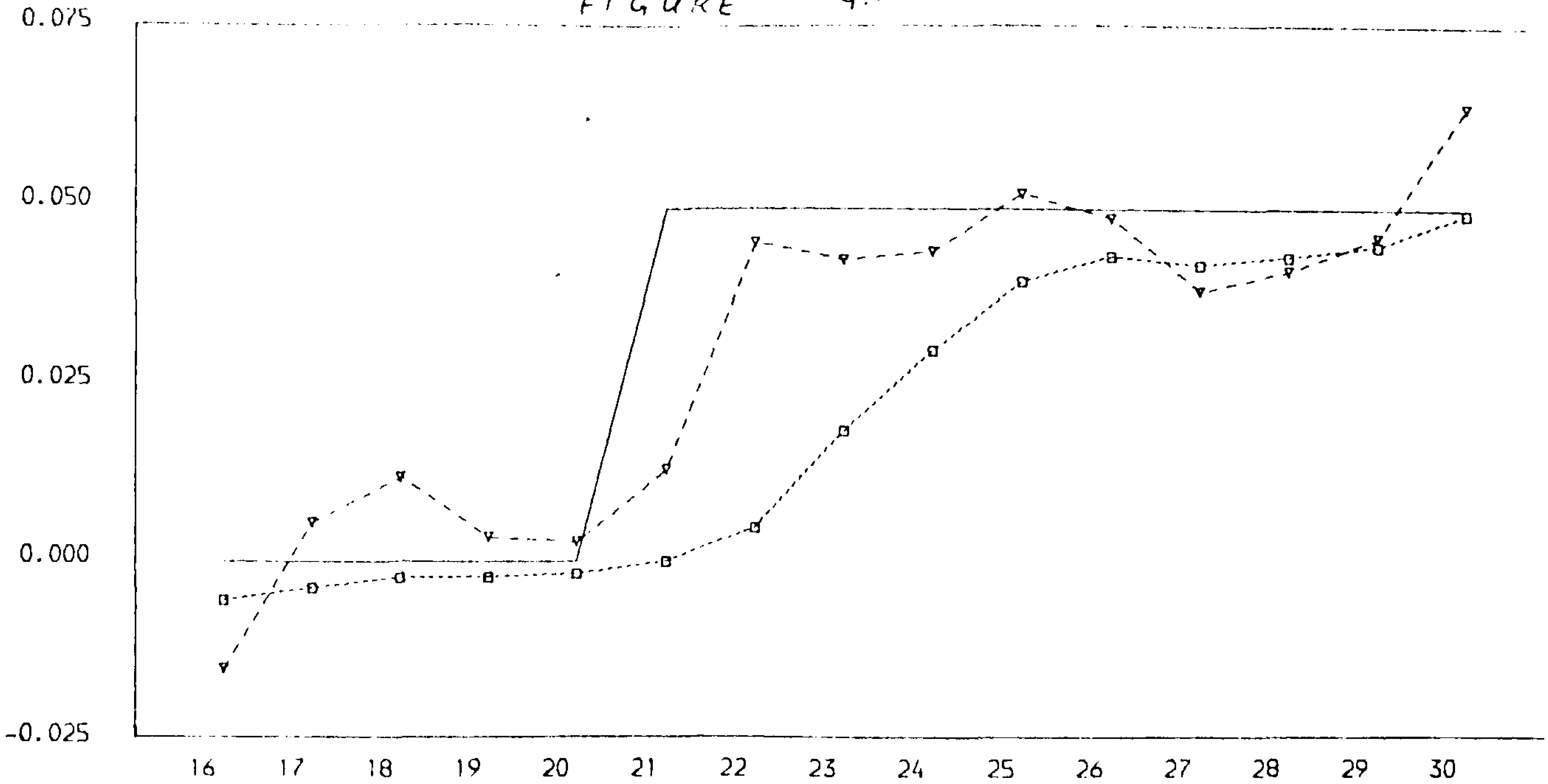
REALISATION 2

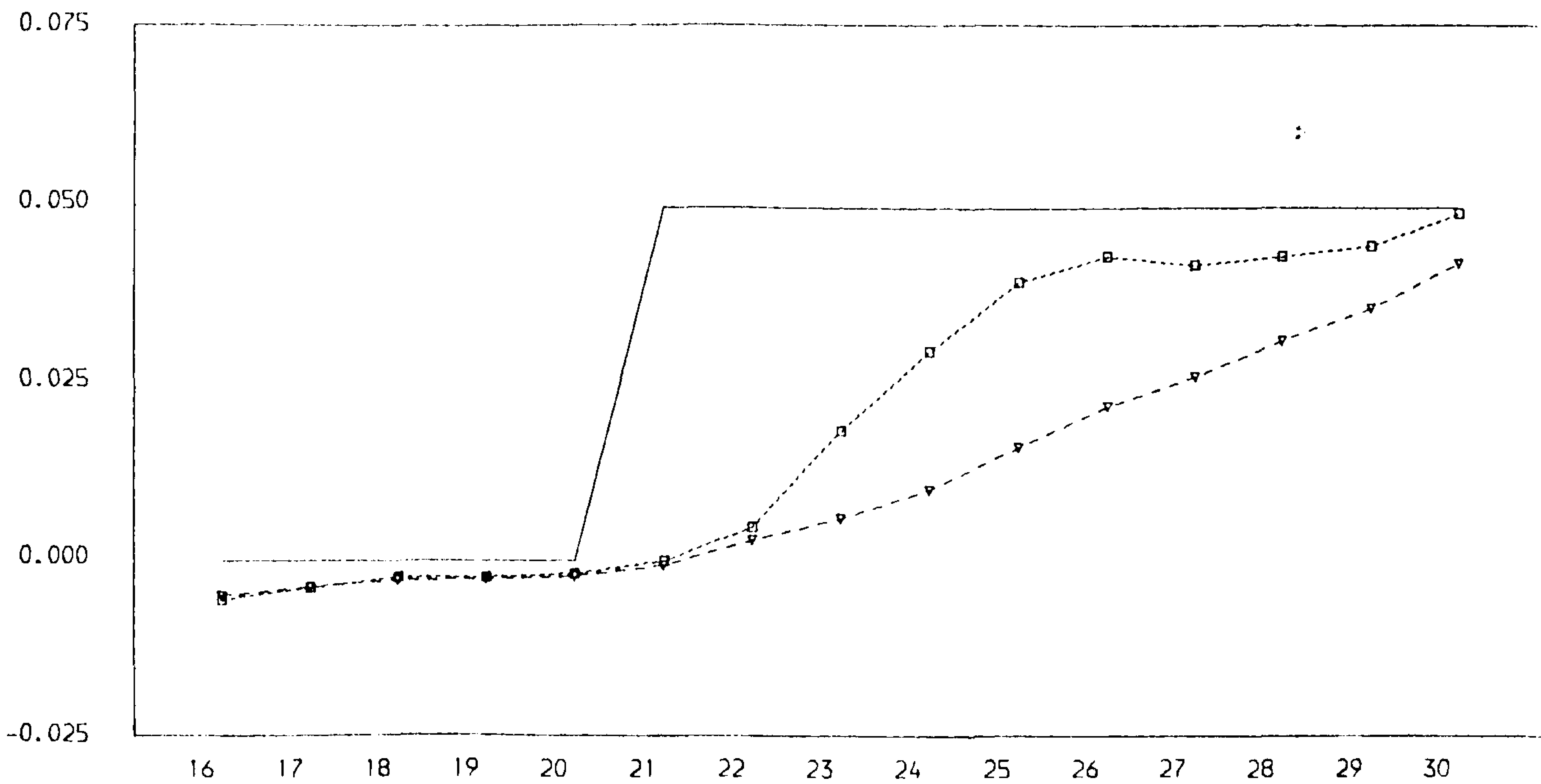
FIGURE G.1

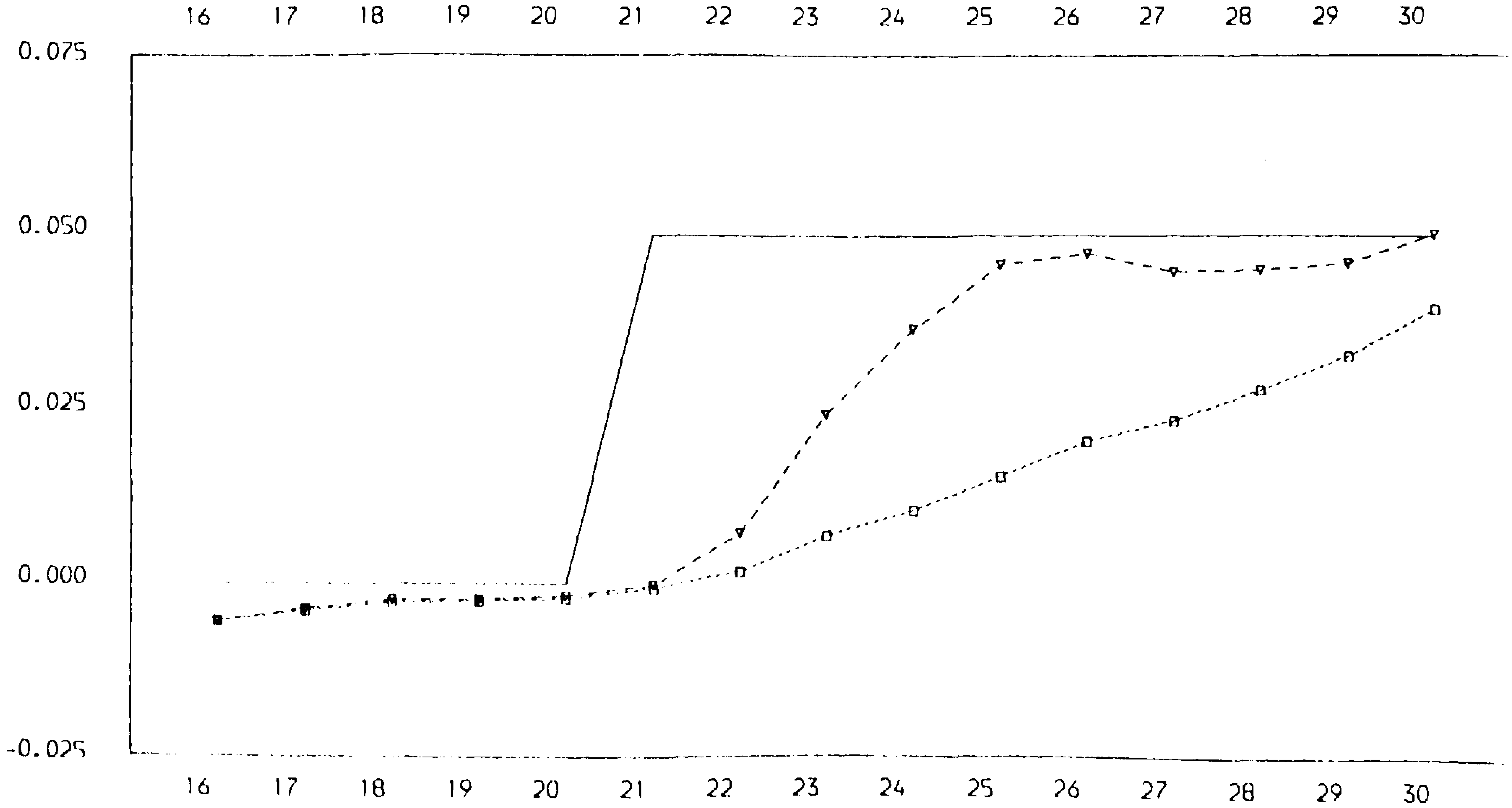


REALISATION 3

FIGURE G.1



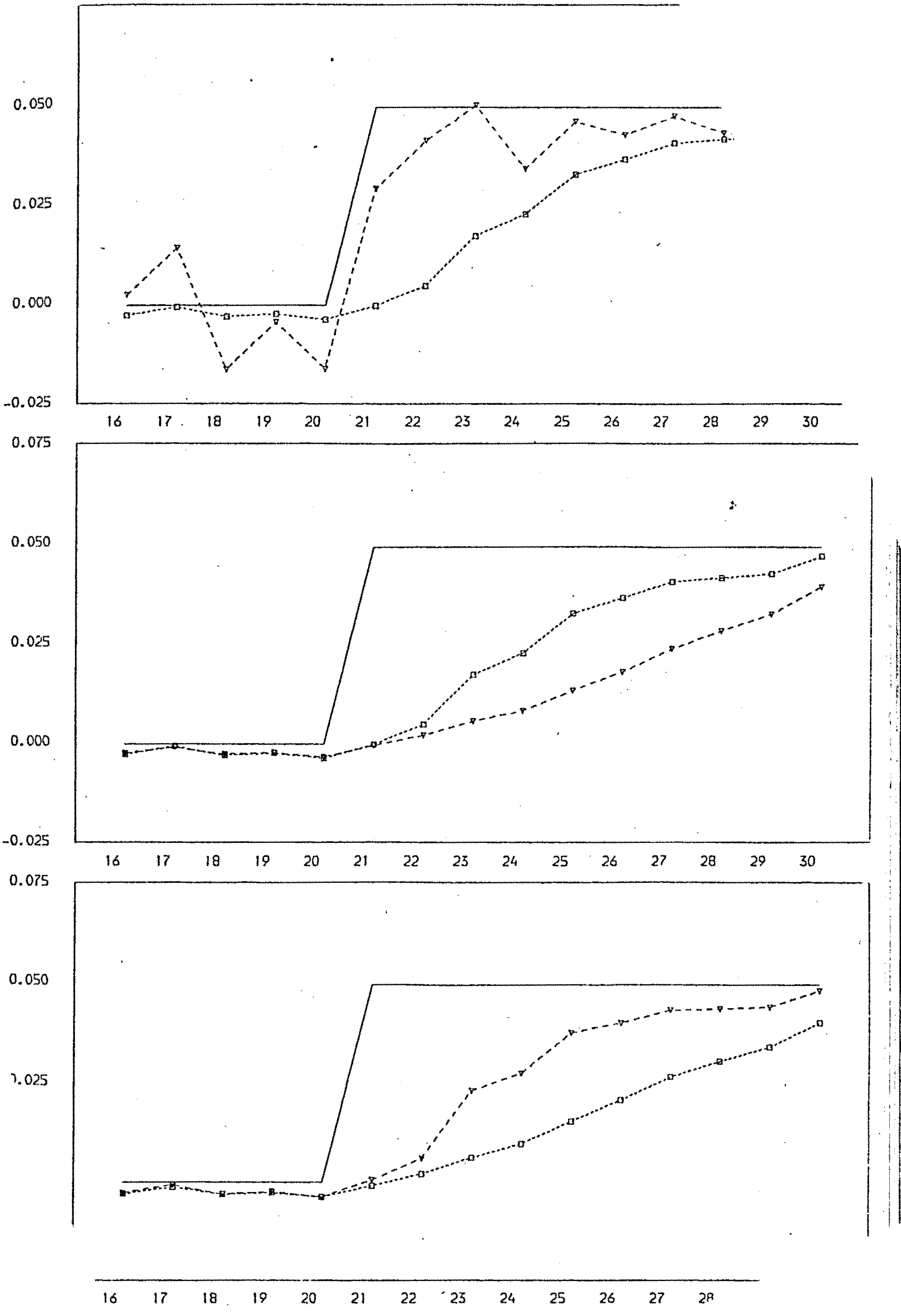




REALISATION 4

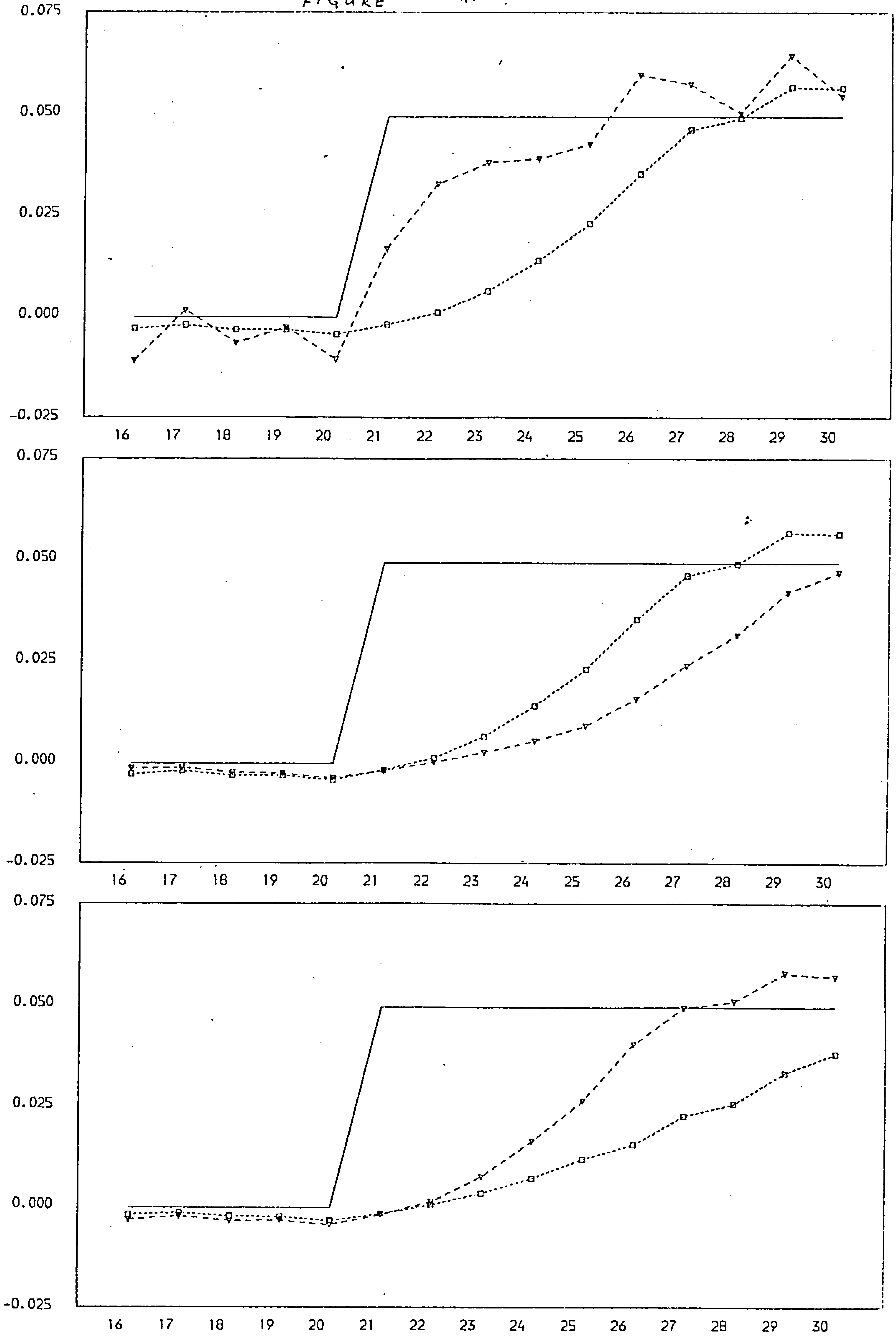
APPENDIX G cont.

FIGURE G.1



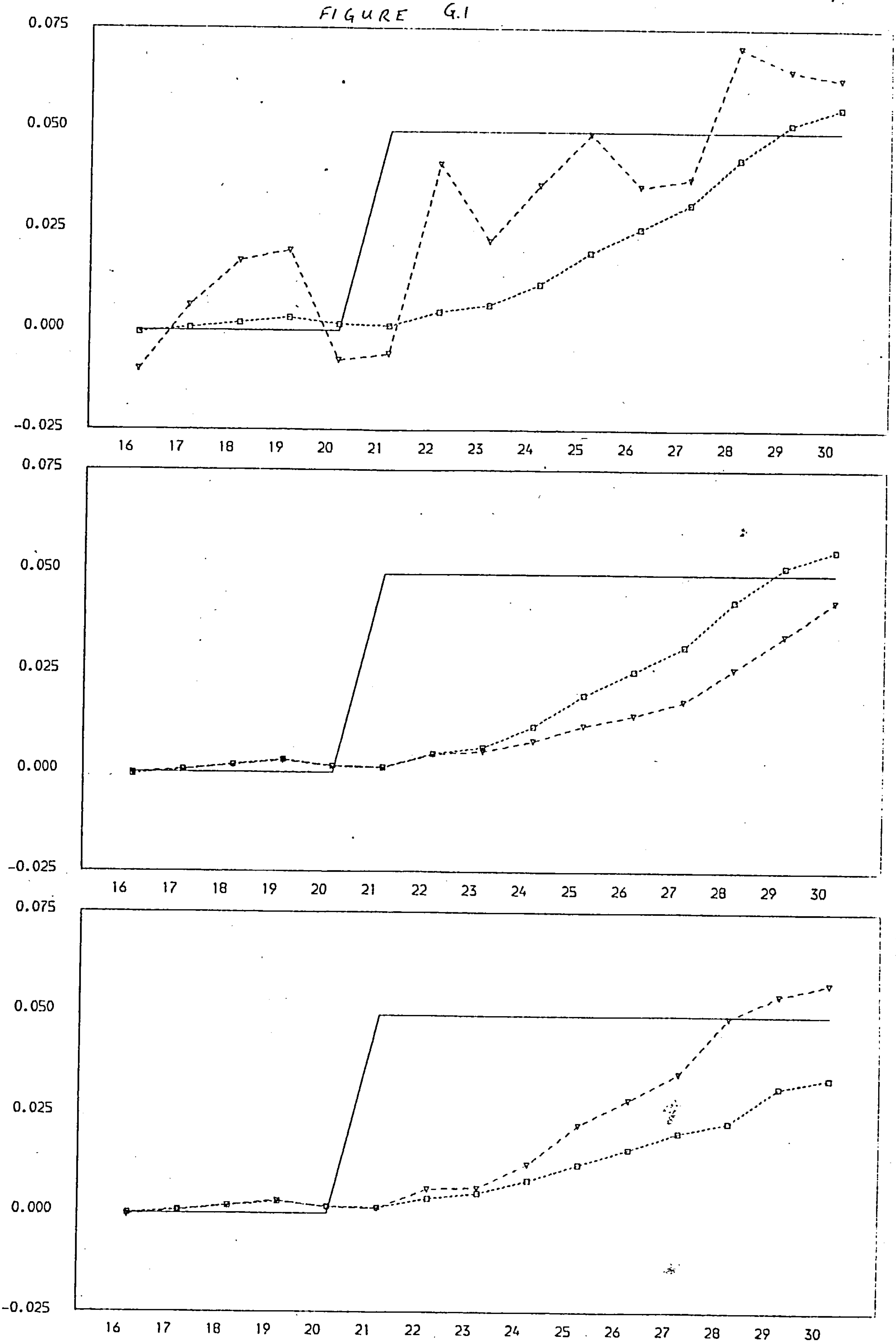
REALISATION 5

FIGURE G.1



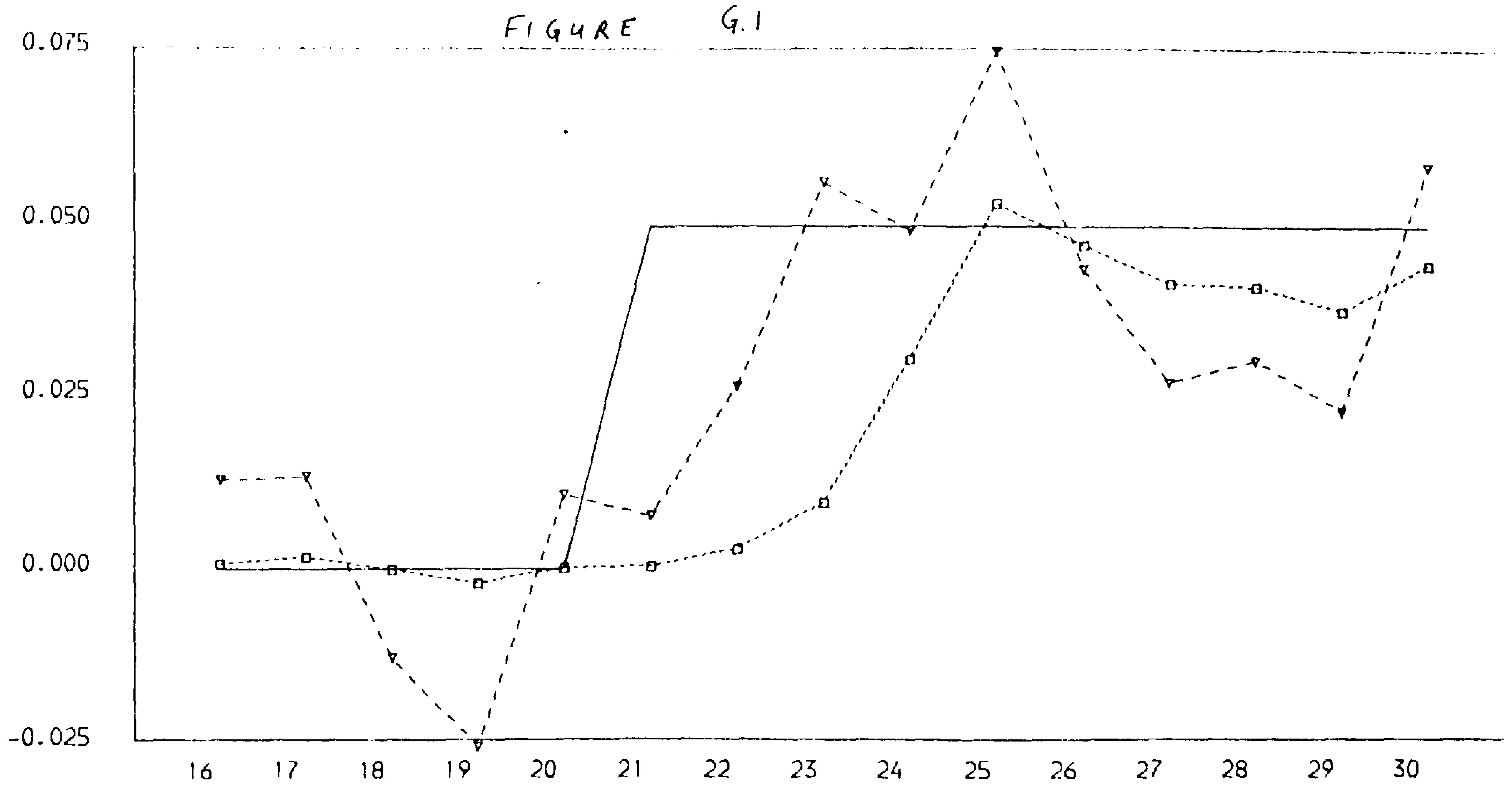
REALISATION 6

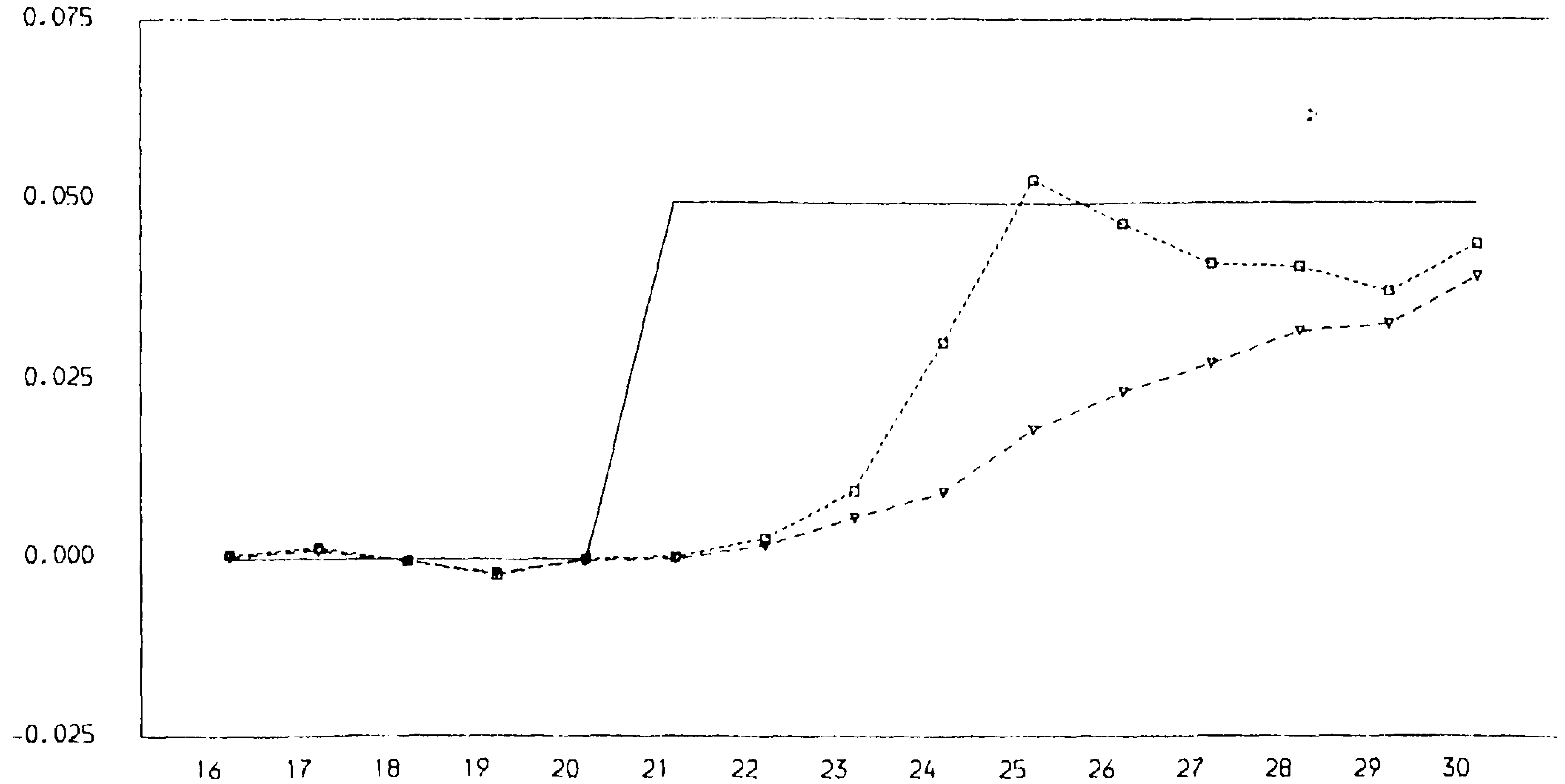
FIGURE G.1

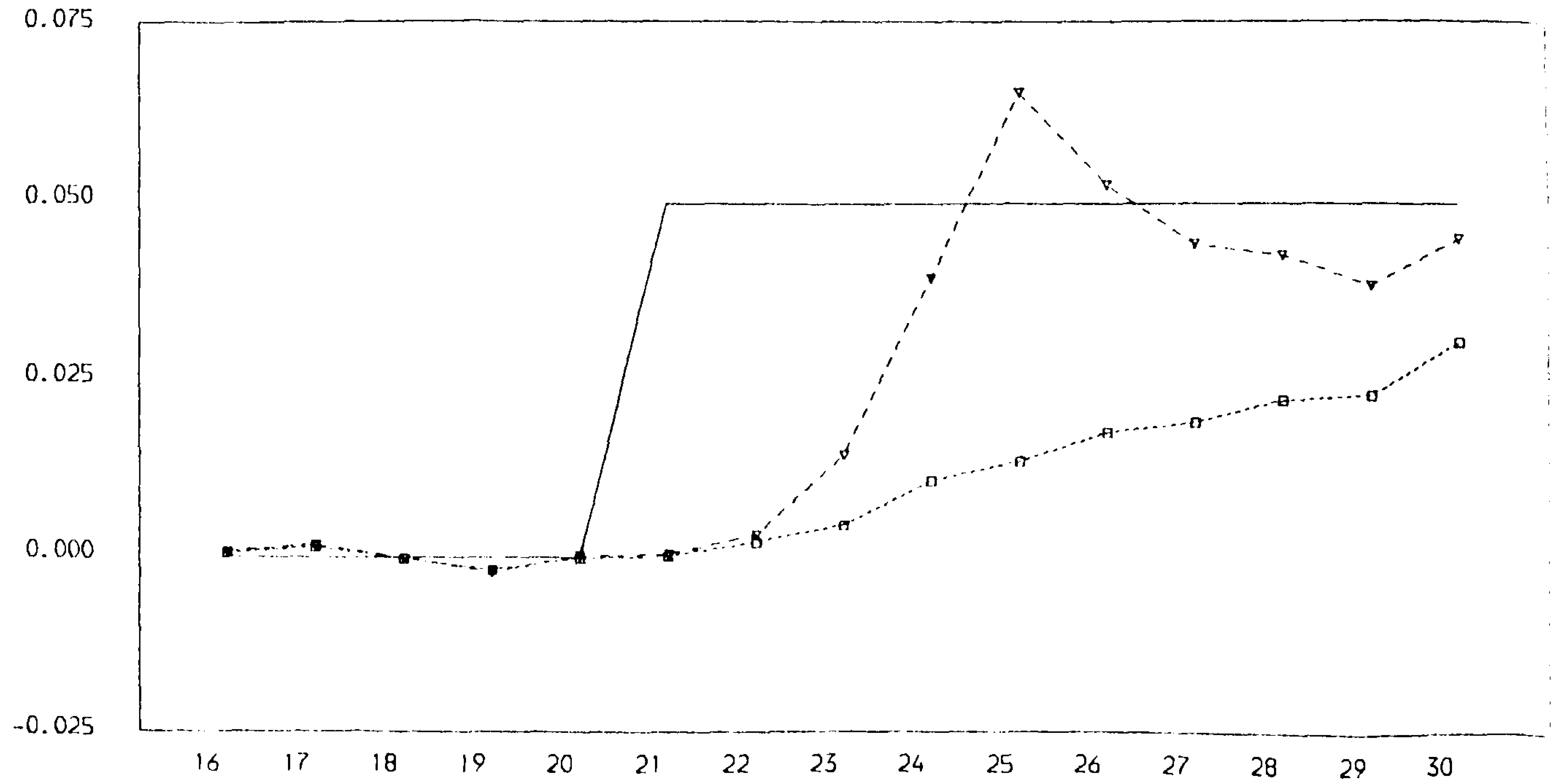


REALISATION 7

FIGURE G.1

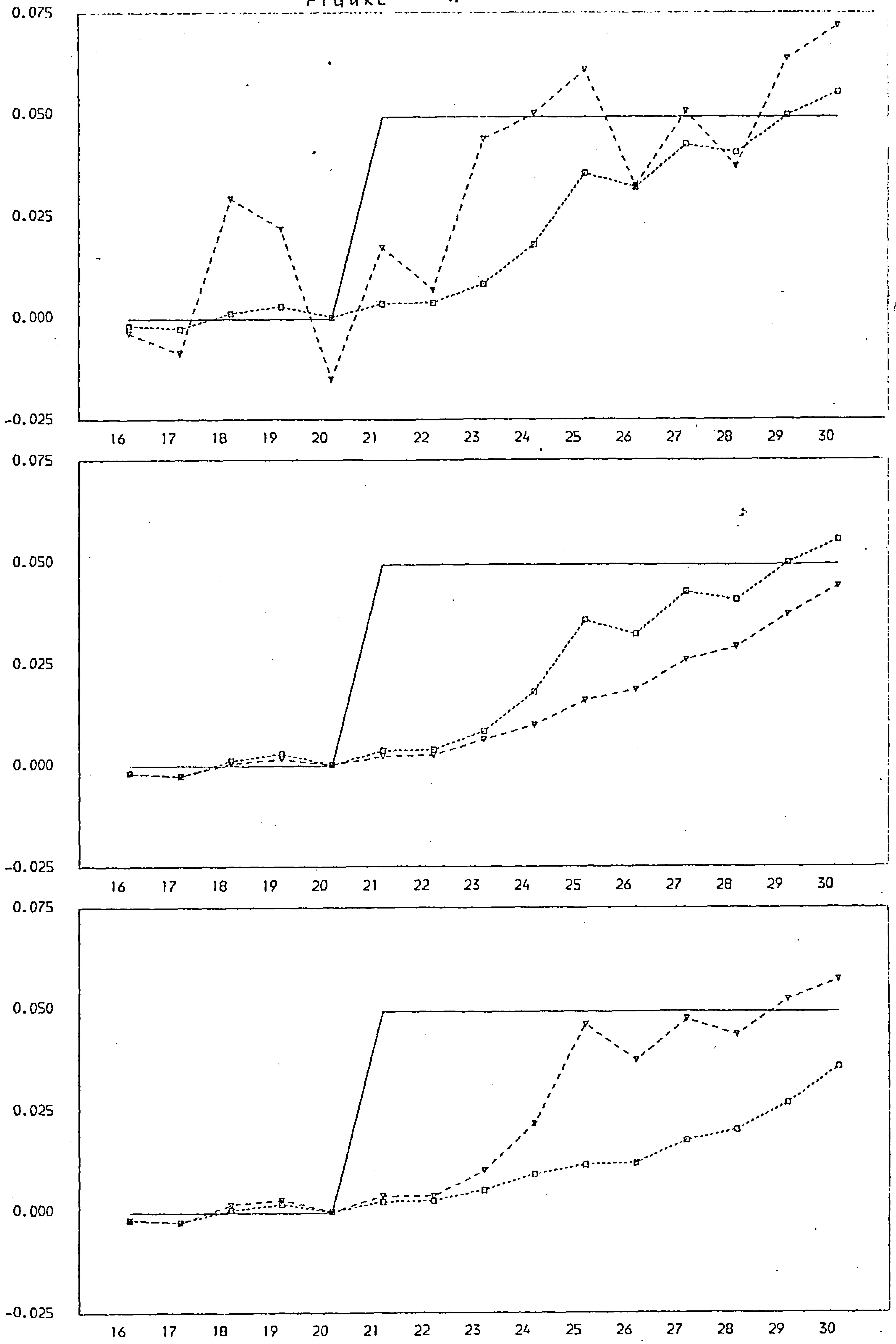






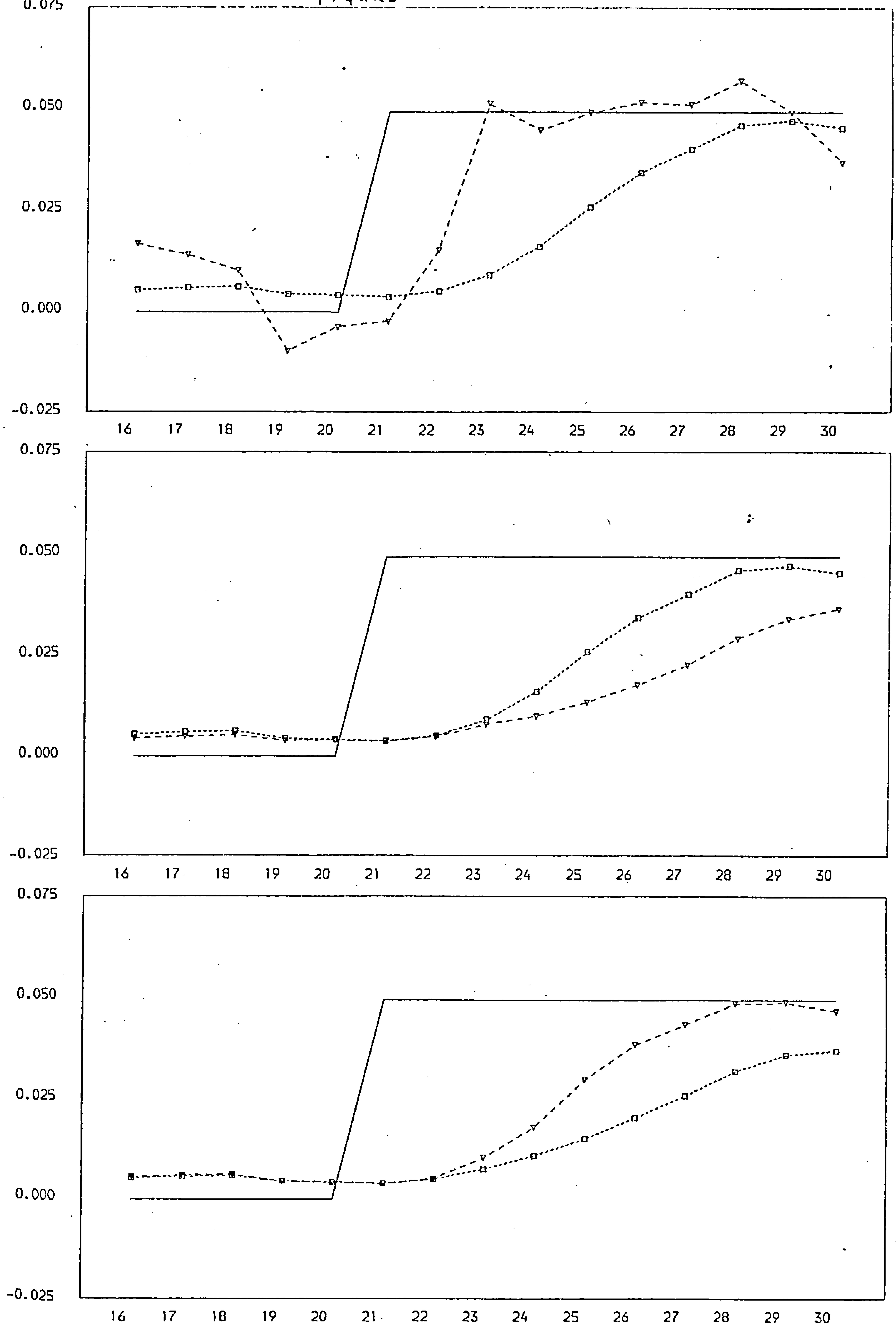
REALISATION 8

FIGURE G.1



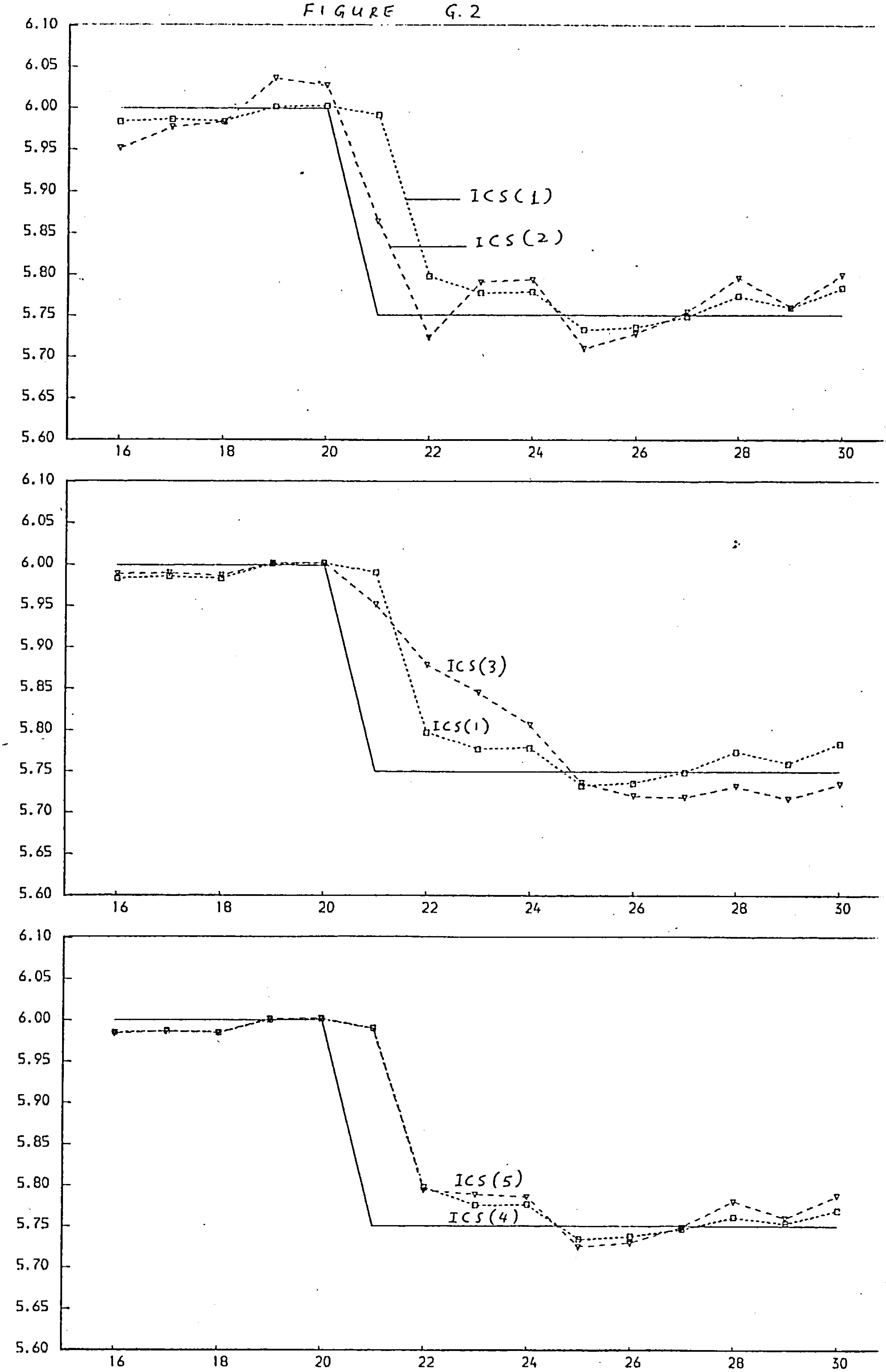
REALISATION 9

FIGURE G.1



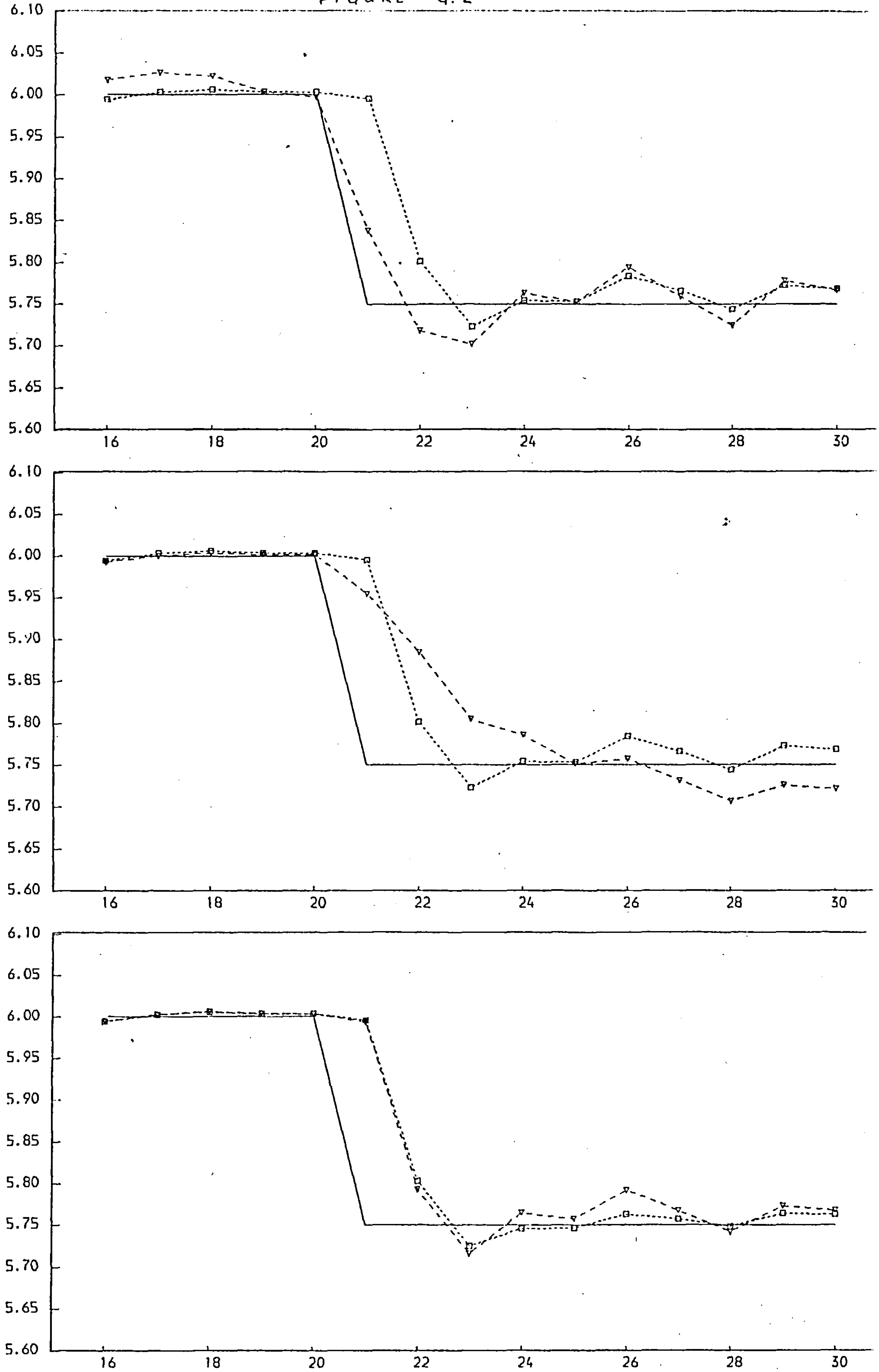
REALISATION 10

FIGURE G.2



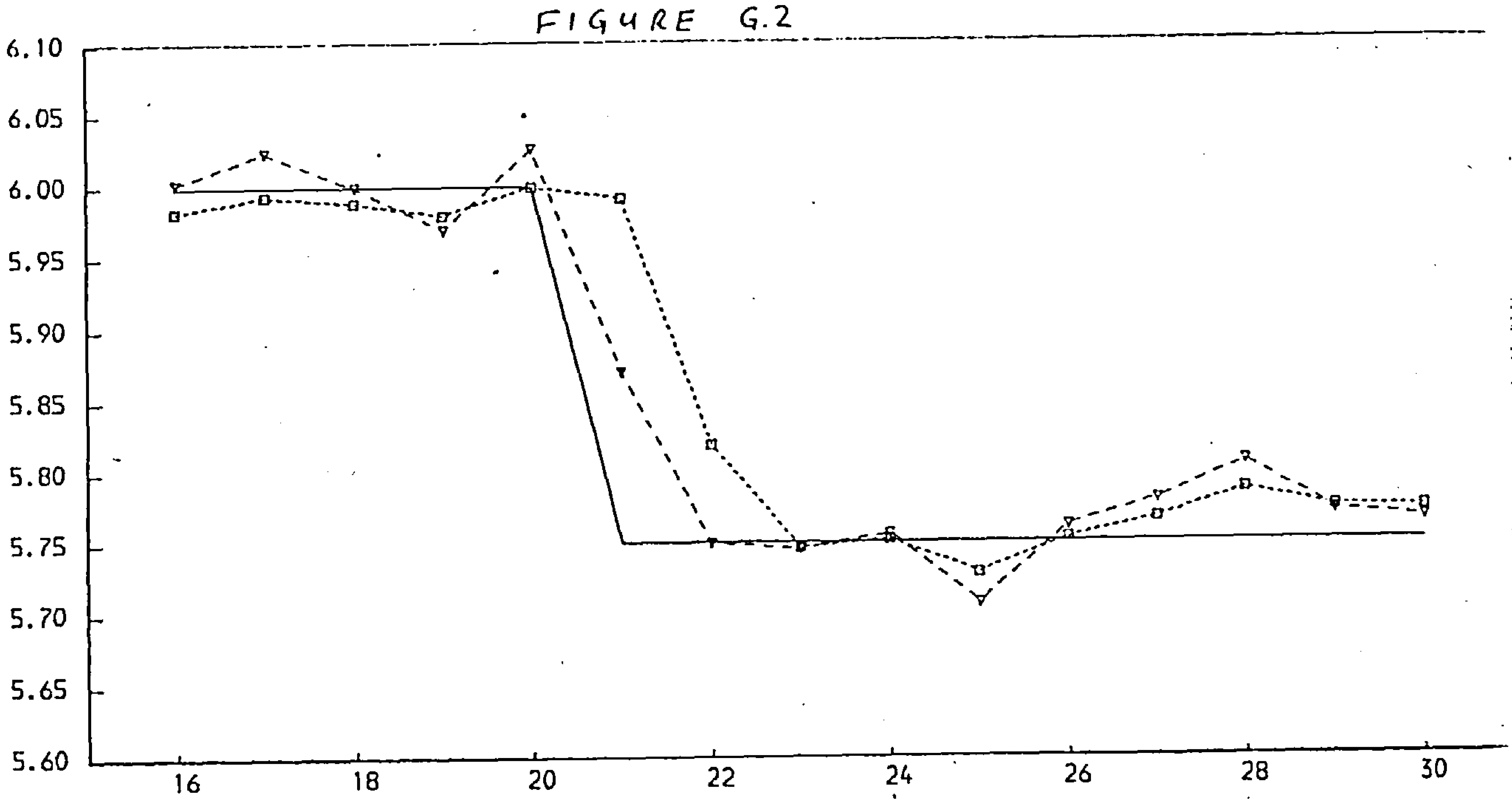
REALISATION 1

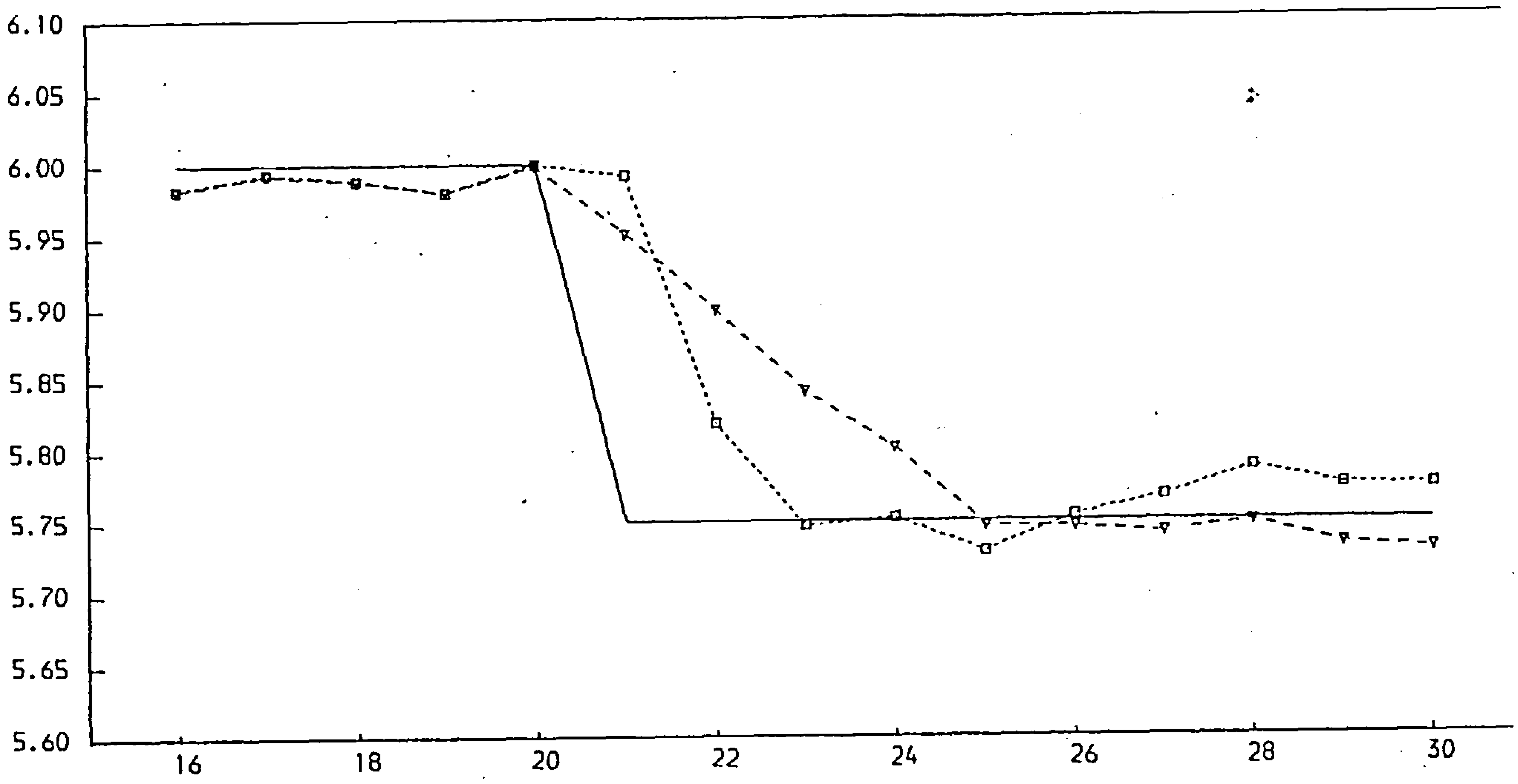
FIGURE G.2

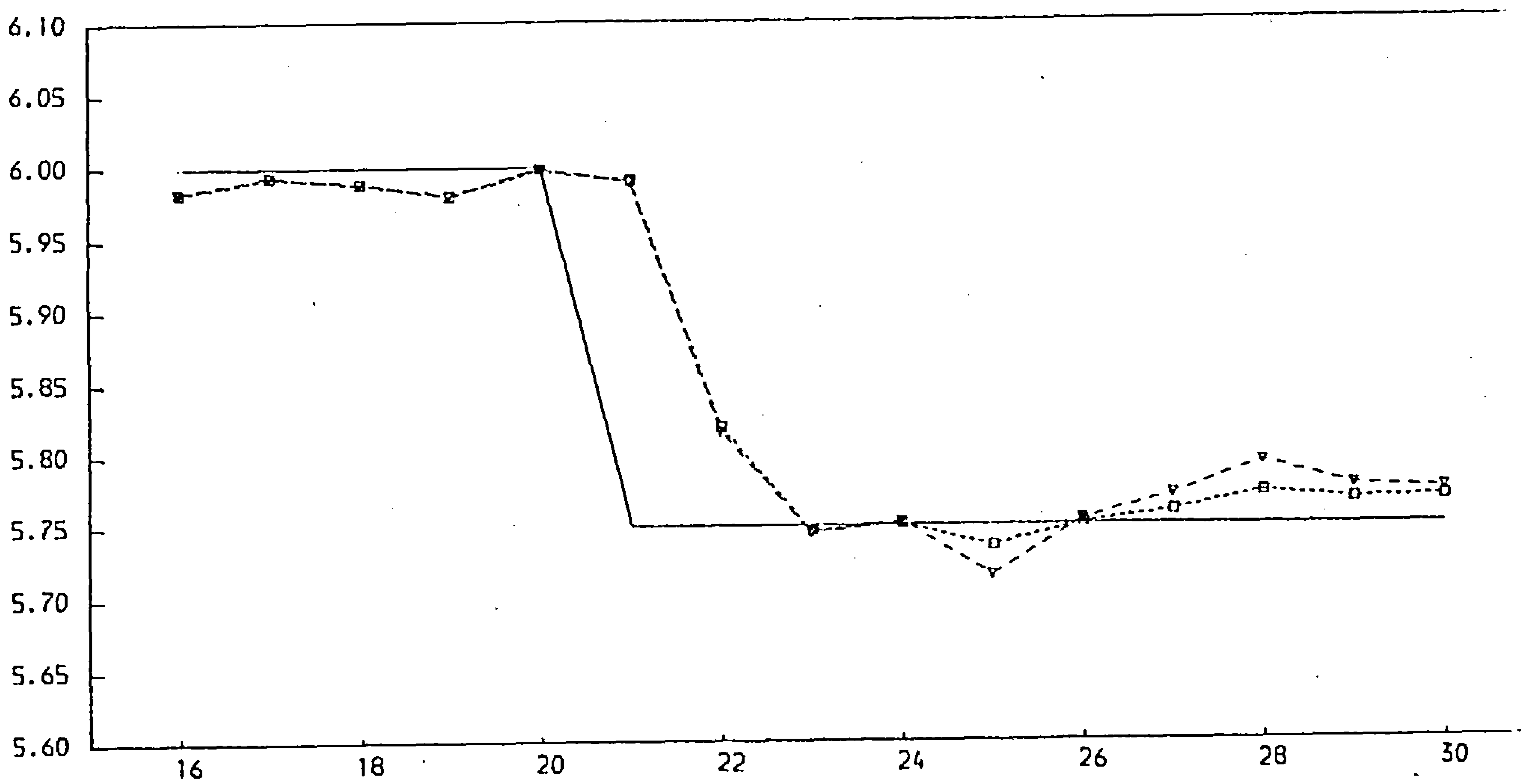


REALISATION 2

FIGURE G.2

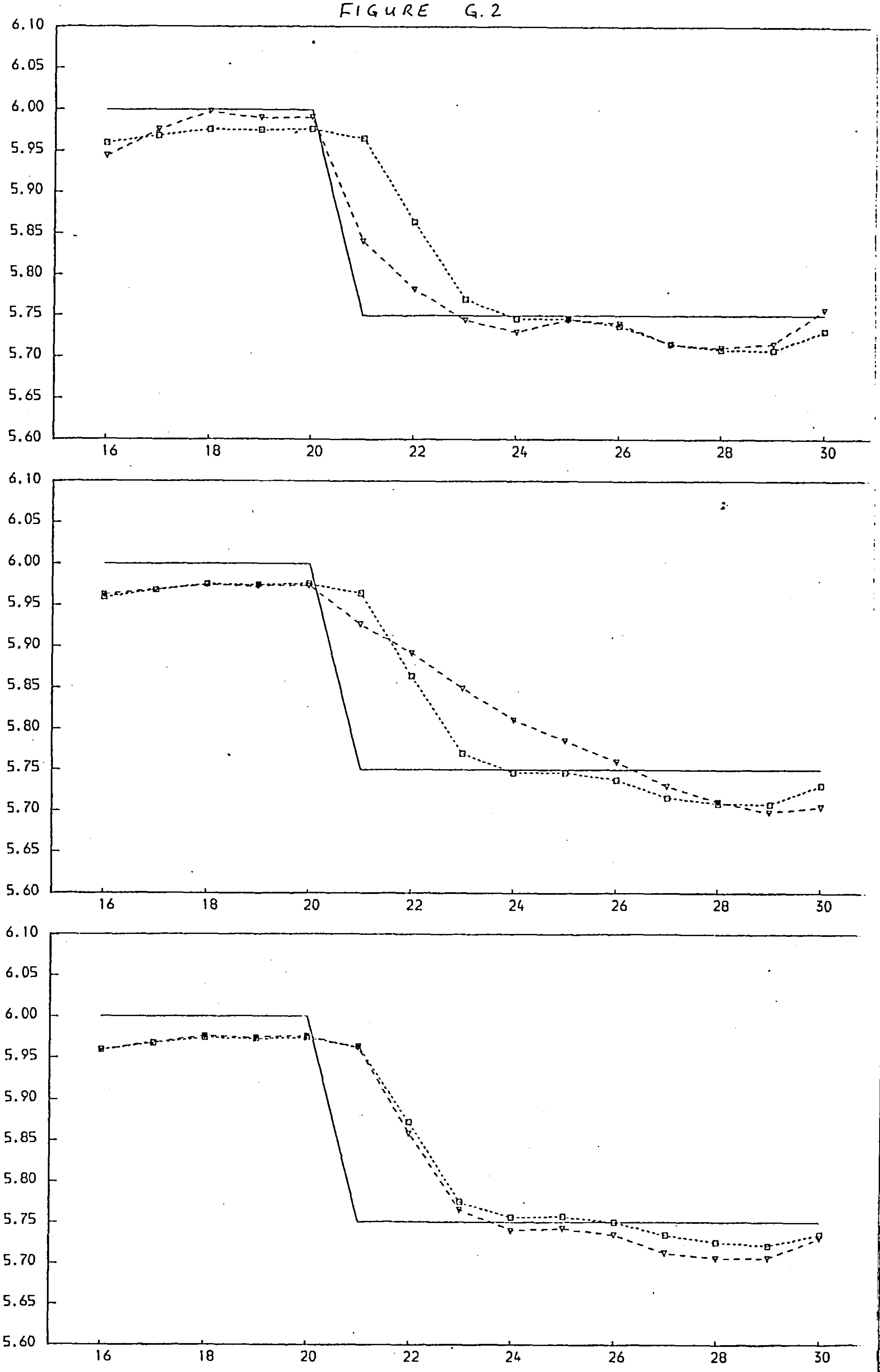






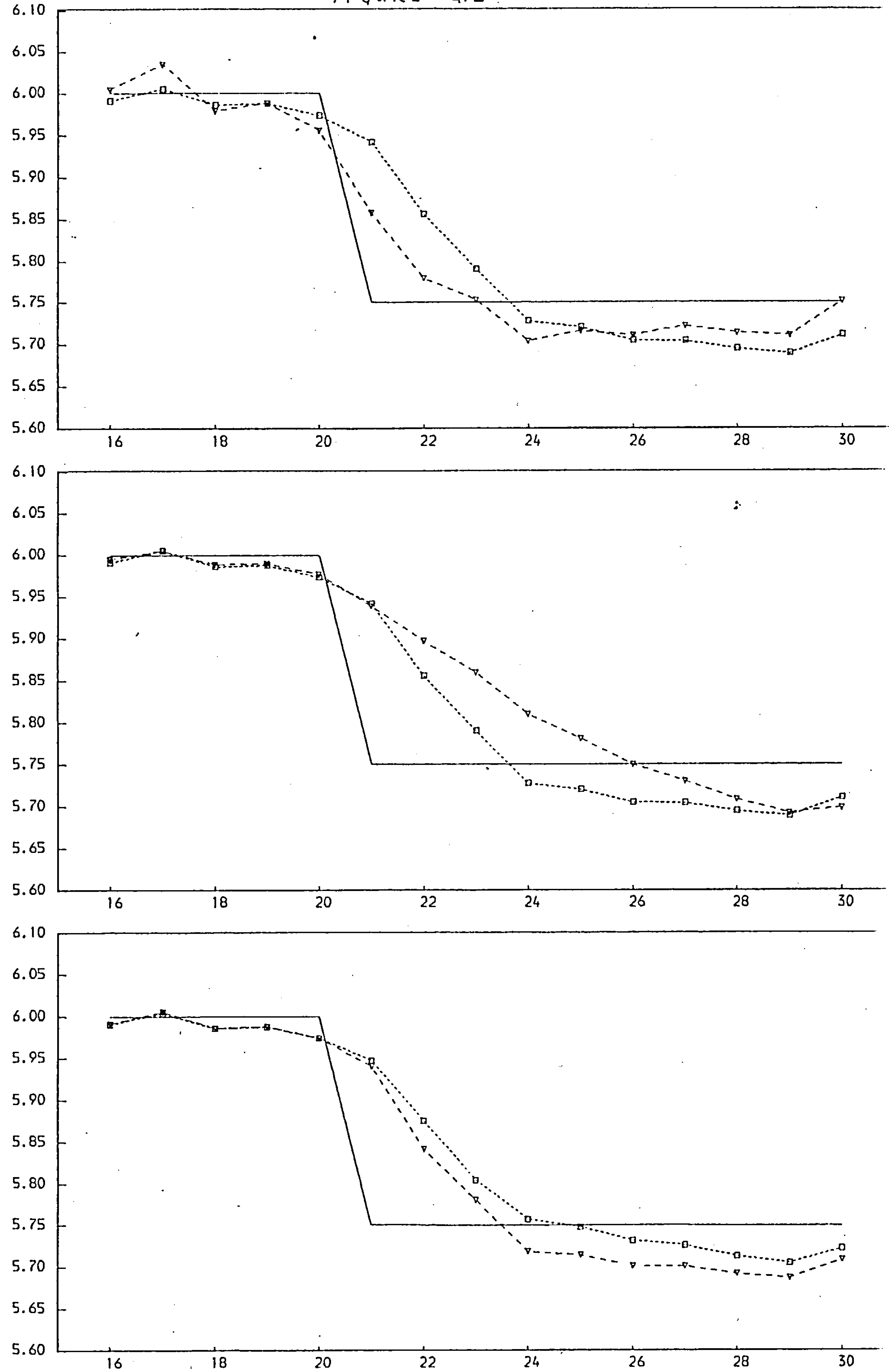
REALISATION 3

FIGURE G.2



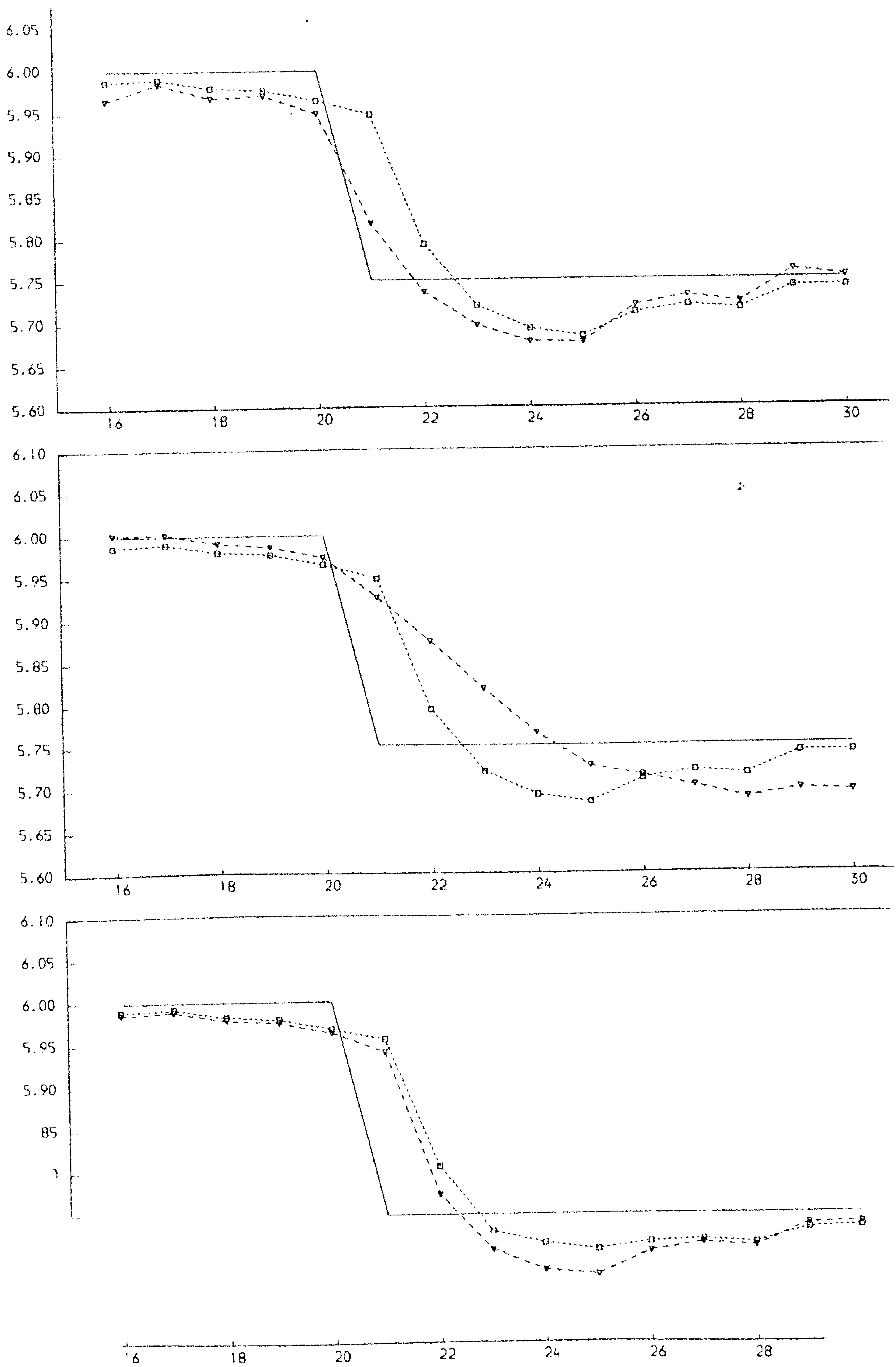
REALISATION 4

FIGURE G.2



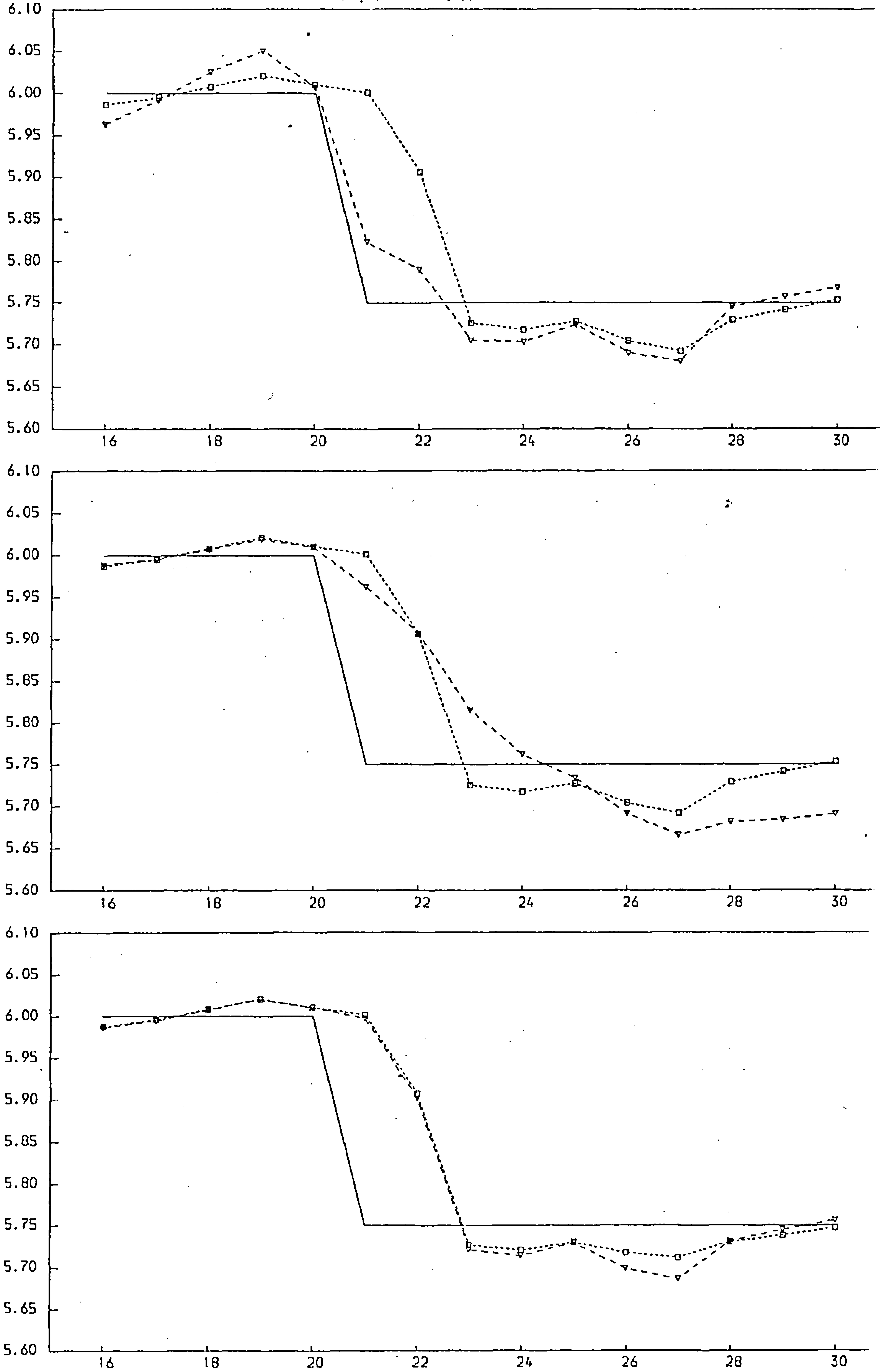
REALISATION 5

FIGURE G.2



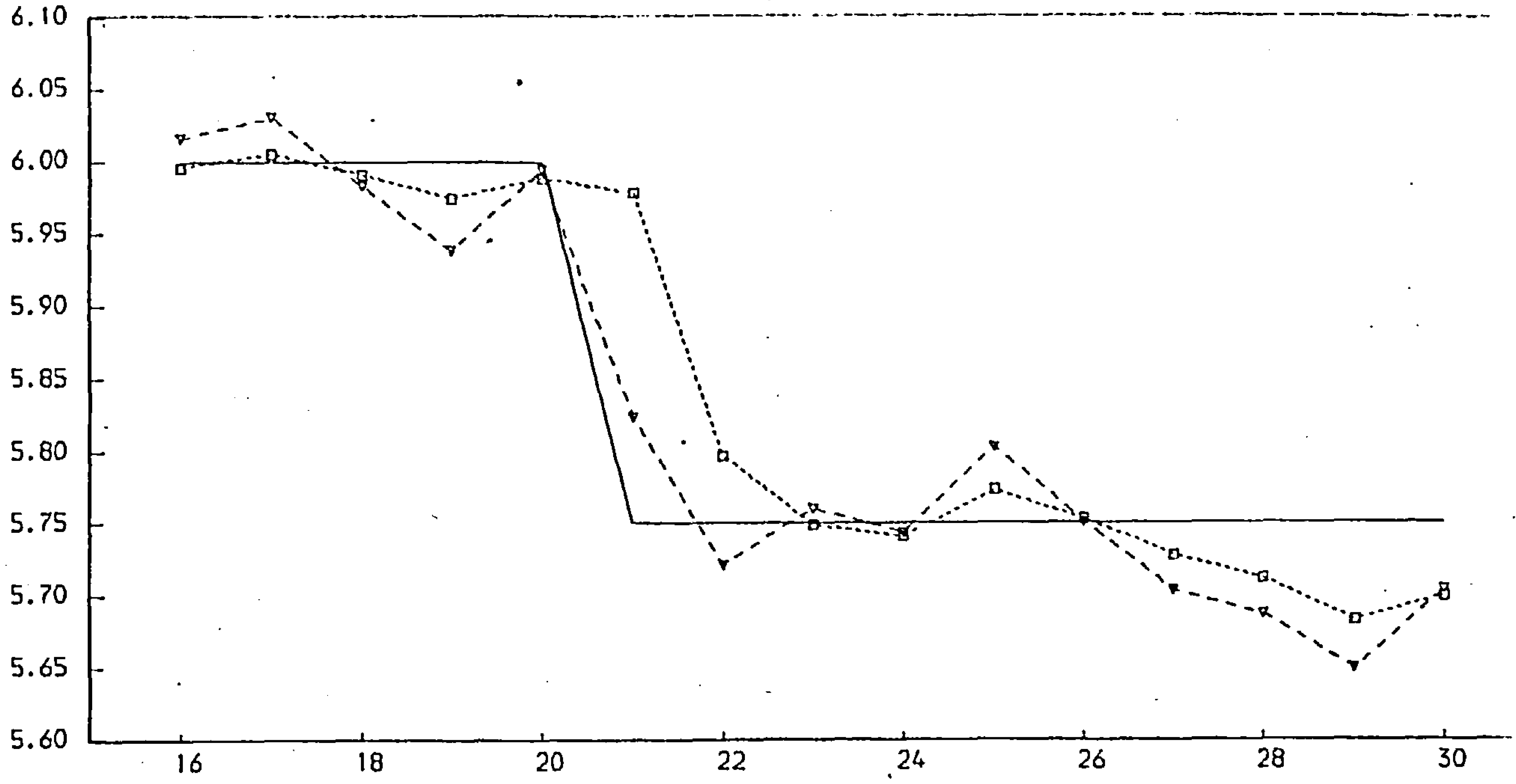
REALISATION 6

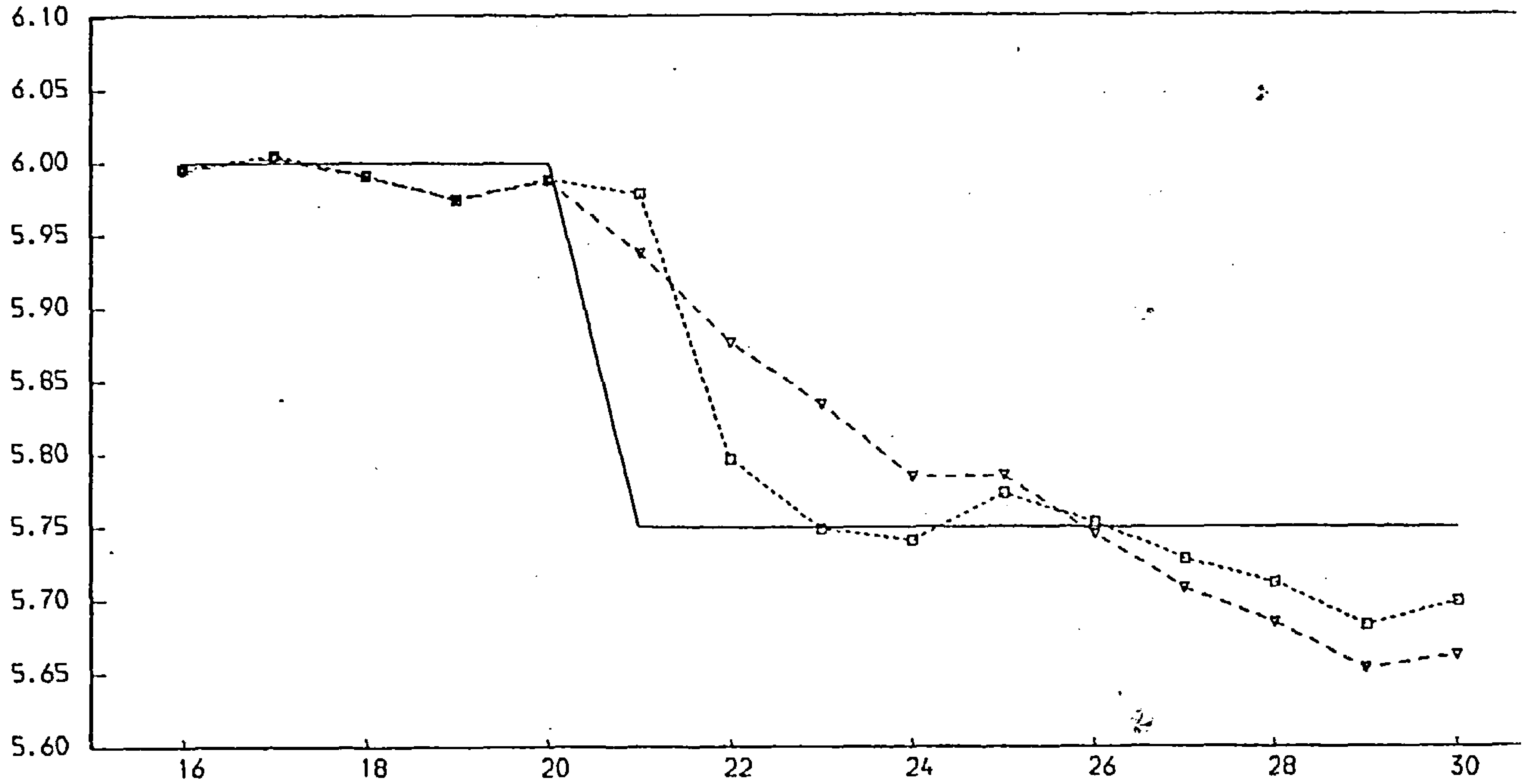
FIGURE G.2

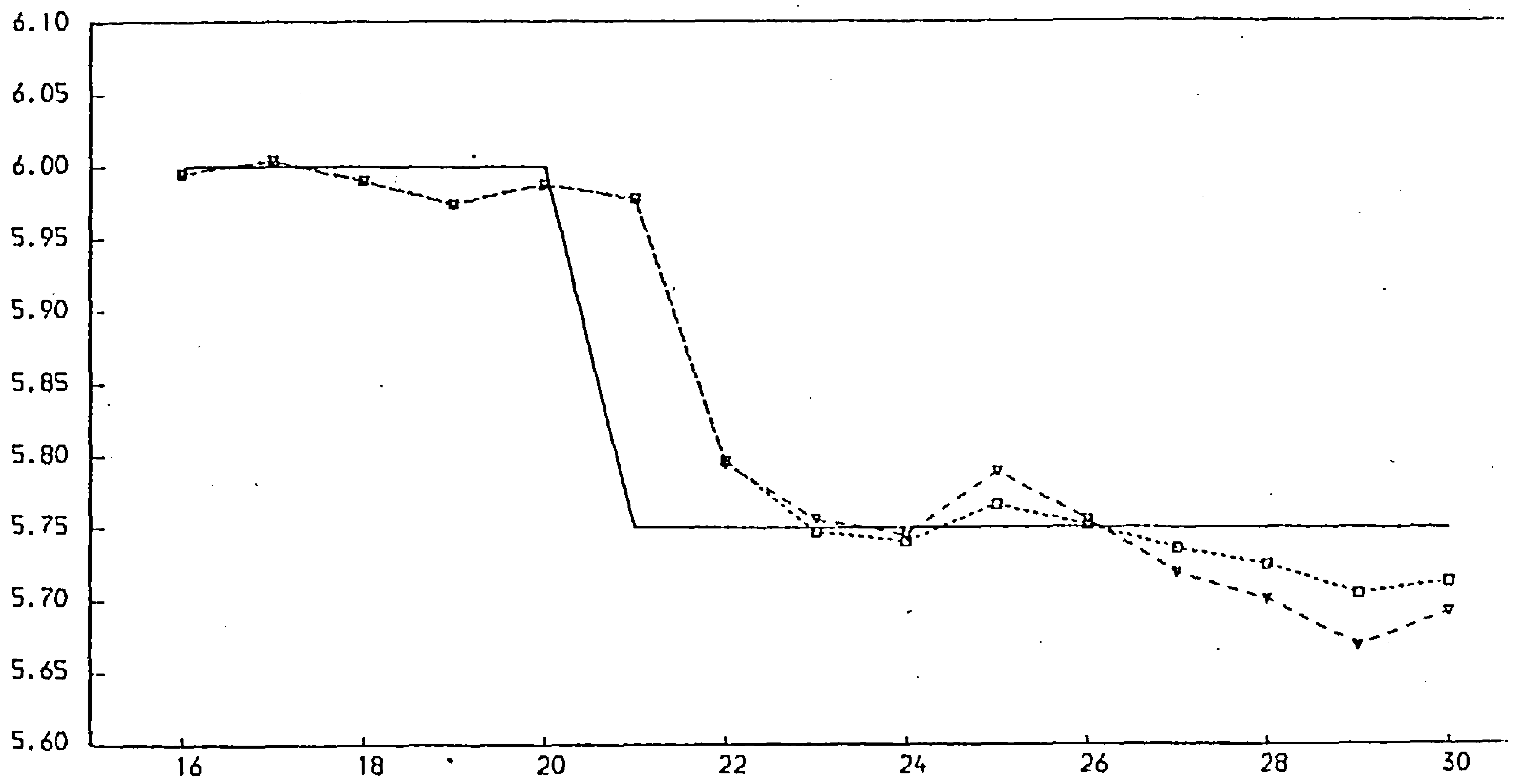


REALISATION 7

FIGURE G.2

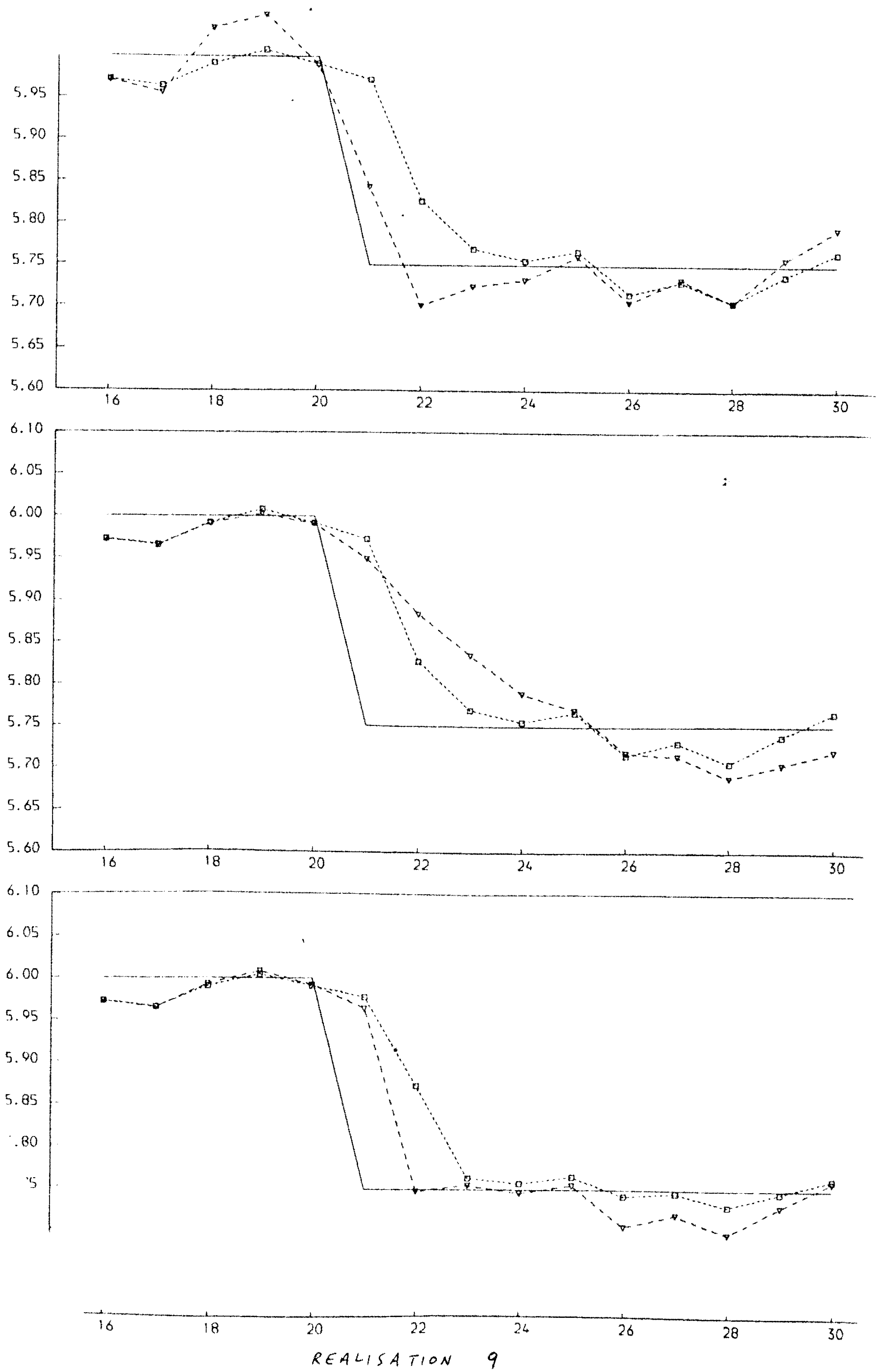






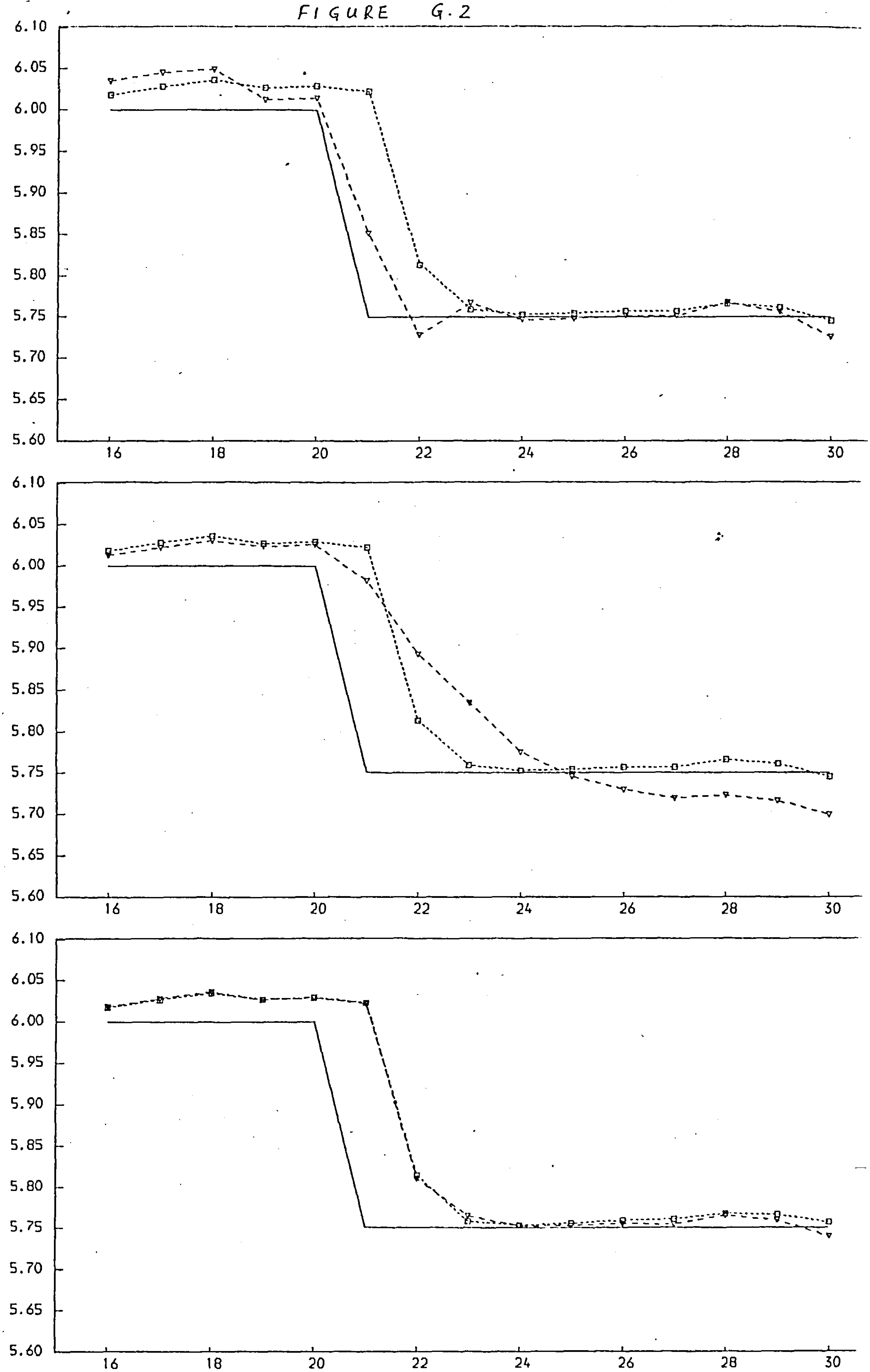
REALISATION 8

FIGURE G.2



REALISATION 9

FIGURE G.2



REALISATION 10

APPENDIX H

This Appendix contains a set of computer program listings corresponding to the following:

(i) Fixed Range Method for the MSM (FRM/MSM)

(ii) Fixed Range Method for the SSM (FRM/SSM)

(iii) Variable Range Method for the SSM (VRM/SSM)

(iv) Class I method (CIM)

(v) Constant variance method (CVM)

The programs have been written in standard Burroughs Basic language as implemented on the B70000/B6000 series of computer systems.

The listings contain a number of comments which together with the theoretical aspects described in chapters 3, 6, and 7, should be sufficient for a good understanding of the programs. However, additional documentation is given below for the most complicated of the programs, FRM/MSM.

FRM/MSM is operating on a large matrix called I containing all the available information at any point in time t. Matrix I is shown in figure H.1 and has 16 rows each corresponding to one of the 16 $i \to j$ transitions, i.e. $1 \to 1$, $2 \to 1$, $3 \to 1$, ... $4 \to 4$ Most of its 29 columns correspond to a particular piece of information such as the one step ahead forecast $\hat{y}$. A detailed explanation of the columns of I are given below:

Column 1 : Each element in this column is either a zero or a one. A zero is interpreted by the program to imply that the associated $i \to j$ transition is excluded. In figure H.1 column 1 is full of 1's and therefore all 16 transitions are considered.

Columns 2, 3 : These correspond to transitions $i \to j$ so that if in a particular row of I, columns 2 and 3 contain, say, 1 and 4 respectively then this implies that this row corresponds to a transition $1 \to 4$. In figure H.1 this occurs in row 13.

Columns 4, 5, 6 : These correspond to $V_{\varepsilon}^{(j)}$, $V_{\mu}^{(j)} + V_{\beta}^{(j)}$ and $V_{\beta}^{(j)}$ respectively. $V_{\varepsilon}^{(j)}$, $V_{\mu}^{(j)}$ and $V_{\beta}^{(j)}$ are used by the program as shown by the set of equations (3.3.7) in chapter 3. For example $r_{11}^{(ij)}$ is the sum of some terms plus $(V_{\mu}^{(j)} + V_{\beta}^{(j)})$ This last term is found by the program, in column 5 and the appropriate row for transition $i \to j$.

<u>Columns 7, 8:</u> These correspond to $\Pi^{(j)}$ and $p_{t-1}^{(i)}$ respectively where $p_{t-1}^{(i)}$ is as defined by (3.3.5)

<u>Columns 9, 10, 11, 12, 13</u> : These correspond to $m_{t-1}^{(i)}$, $b_{t-1}^{(i)}$ $c_{11,\ t-1}^{(i)}$, $c_{12,\ t-1}^{(i)}$ and $c_{22,\ t-1}^{(i)}$ respectively, as defined by (3.3.4). Consider for example the equation for $r_{12}^{(ij)}$ from (3.3.7) and for a transition $3 \rightarrow 4$:

$$r_{12}^{(34)} = c_{12,\ t-1}^{(3)} + c_{22,\ t-1}^{(3)} + V_{\beta}^{(4)}$$

This can be written as follows using elements from the matrix I and noting that transition $3 \rightarrow 4$ corresponds to row 15 of I :

$$r_{12}^{(34)} = I\ (15,\ 12) + I\ (15,\ 13) + I\ (15,\ 6) \qquad \text{(H.1)}$$

<u>Columns 14 to 29</u>: The program operates on columns 4 to 13 (prior information) to produce posterior information given y_t (the latest observation) and stores them in columns 14 to 29. These correspond to the following variables as defined in the set of equations (3.3.7), (3.3.8), (3.3.9) and (3.3.12):

$\hat{y}^{(ij)}$, $e^{(ij)}$, $r_{11}^{(ij)}$, $r_{12}^{(ij)}$, $r_{22}^{(ij)}$, $\hat{Y}^{(ij)}$, $A_1^{(ij)}$ $A_2^{(ij)}$, $m^{(ij)}$, $b^{(ij)}$, $c_{11}^{(ij)}$, $c_{12}^{(ij)}$, $c_{22}^{(ij)}$, $L^{(ij)}$, $p^{(ij)}$ and $p_t^{(j)}$ respectively.

This together with the comments on the actual program listing completes the documentation of FRM/MSM.

MATRIX I

	1	2	3	4	5	6	7	8	9	10	11	12	23	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1	1	1	1	$V_\varepsilon^{(1)}$	$V_\mu^{(1)}+V_\beta^{(1)}$	$V_\beta^{(1)}$	$\pi^{(1)}$	$P_{t-1}^{(1)}$	$m_{t-1}^{(1)}$	$b_{t-1}^{(1)}$	$c_{11,t-1}$	$c_{12,t-1}$	$c_{22,t-1}$															$\rho^{(11)}$	$\rho_t^{(1)}$
2	1	2	1	$V_\varepsilon^{(1)}$	$V_\mu^{(1)}+V_\beta^{(1)}$	$V_\beta^{(1)}$	$\pi^{(1)}$	$P_{t-1}^{(2)}$	$m_{t-1}^{(2)}$	$b_{t-1}^{(2)}$	$c_{11,t-1}$	$c_{12,t-1}$	$c_{22,t-1}$															$\rho^{(21)}$	$\rho_t^{(1)}$
3	1	3	1	$V_\varepsilon^{(1)}$	$V_\mu^{(1)}+V_\beta^{(1)}$	$V_\beta^{(1)}$	$\pi^{(1)}$	$P_{t-1}^{(3)}$	$m_{t-1}^{(3)}$	$b_{t-1}^{(3)}$	$c_{11,t-1}$	$c_{12,t-1}$	$c_{22,t-1}$															$\rho^{(31)}$	$\rho_t^{(1)}$
4	1	4	1	$V_\varepsilon^{(1)}$	$V_\mu^{(1)}+V_\beta^{(1)}$	$V_\beta^{(1)}$	$\pi^{(1)}$	$P_{t-1}^{(4)}$	$m_{t-1}^{(4)}$	$b_{t-1}^{(4)}$	$c_{11,t-1}$	$c_{12,t-1}$	$c_{22,t-1}$															$\rho^{(41)}$	$\rho_t^{(1)}$
5	1	1	2	$V_\varepsilon^{(2)}$	$V_\mu^{(2)}+V_\beta^{(2)}$	$V_\beta^{(2)}$	$\pi^{(2)}$	$P_{t-1}^{(1)}$	$m_{t-1}^{(1)}$	$b_{t-1}^{(1)}$																		$\rho^{(12)}$	$\rho_t^{(2)}$
6	1	2	2	$V_\varepsilon^{(2)}$	$V_\mu^{(2)}+V_\beta^{(2)}$	$V_\beta^{(2)}$	$\pi^{(2)}$	$P_{t-1}^{(2)}$	$m_{t-1}^{(2)}$	$b_{t-1}^{(2)}$																		$\rho^{(22)}$	$\rho_t^{(2)}$
7	1	3	2	$V_\varepsilon^{(2)}$	$V_\mu^{(2)}+V_\beta^{(2)}$	$V_\beta^{(2)}$	$\pi^{(2)}$	$P_{t-1}^{(3)}$	$m_{t-1}^{(3)}$	$b_{t-1}^{(3)}$																		$\rho^{(32)}$	$\rho_t^{(2)}$
8	1	4	2	$V_\varepsilon^{(2)}$	$V_\mu^{(2)}+V_\beta^{(2)}$	$V_\beta^{(2)}$	$\pi^{(2)}$	$P_{t-1}^{(4)}$	$m_{t-1}^{(4)}$	$b_{t-1}^{(4)}$																		$\rho^{(42)}$	$\rho_t^{(2)}$
9	1	1	3	$V_\varepsilon^{(3)}$	$V_\mu^{(3)}+V_\beta^{(3)}$	$V_\beta^{(3)}$	$\pi^{(3)}$	$P_{t-1}^{(1)}$	$m_{t-1}^{(1)}$	$b_{t-1}^{(1)}$		e.t.c								e.t.c								$\rho^{(13)}$	$\rho_t^{(3)}$
10	1	2	3	$V_\varepsilon^{(3)}$	$V_\mu^{(3)}+V_\beta^{(3)}$	$V_\beta^{(3)}$	$\pi^{(3)}$	$P_{t-1}^{(2)}$	$m_{t-1}^{(2)}$	$b_{t-1}^{(2)}$																		$\rho^{(23)}$	$\rho_t^{(3)}$
11	1	3	3	$V_\varepsilon^{(3)}$	$V_\mu^{(3)}+V_\beta^{(3)}$	$V_\beta^{(3)}$	$\pi^{(3)}$	$P_{t-1}^{(3)}$	$m_{t-1}^{(3)}$	$b_{t-1}^{(3)}$																		$\rho^{(33)}$	$\rho_t^{(3)}$
12	1	4	3	$V_\varepsilon^{(3)}$	$V_\mu^{(3)}+V_\beta^{(3)}$	$V_\beta^{(3)}$	$\pi^{(3)}$	$P_{t-1}^{(4)}$	$m_{t-1}^{(4)}$	$b_{t-1}^{(4)}$																		$\rho^{(43)}$	$\rho_t^{(3)}$
13	1	1	4	$V_\varepsilon^{(4)}$	$V_\mu^{(4)}+V_\beta^{(4)}$	$V_\beta^{(4)}$	$\pi^{(4)}$	$P_{t-1}^{(1)}$	$m_{t-1}^{(1)}$	$b_{t-1}^{(1)}$																		$\rho^{(14)}$	$\rho_t^{(4)}$
14	1	2	4	$V_\varepsilon^{(4)}$	$V_\mu^{(4)}+V_\beta^{(4)}$	$V_\beta^{(4)}$	$\pi^{(4)}$	$P_{t-1}^{(2)}$	$m_{t-1}^{(2)}$	$b_{t-1}^{(2)}$																		$\rho^{(24)}$	$\rho_t^{(4)}$
15	1	3	4	$V_\varepsilon^{(4)}$	$V_\mu^{(4)}+V_\beta^{(4)}$	$V_\beta^{(4)}$	$\pi^{(4)}$	$P_{t-1}^{(3)}$	$m_{t-1}^{(3)}$	$b_{t-1}^{(3)}$																		$\rho^{(34)}$	$\rho_t^{(4)}$
16	1	4	4	$V_\varepsilon^{(4)}$	$V_\mu^{(4)}+V_\beta^{(4)}$	$V_\beta^{(4)}$	$\pi^{(4)}$	$P_{t-1}^{(4)}$	$m_{t-1}^{(4)}$	$b_{t-1}^{(4)}$																		$\rho^{(44)}$	$\rho_t^{(4)}$
		i	j	$V_\varepsilon^{(j)}$	$V_\mu^{(j)}+V_\beta^{(j)}$	$V_\beta^{(j)}$	$\pi^{(j)}$	$P_{t-1}^{(i)}$	$m_{t-1}^{(i)}$	$b_{t-1}^{(i)}$	$c_{11,t-1}^{(i)}$	$c_{12,t-1}^{(i)}$	$c_{22,t-1}^{(i)}$	$\hat{y}^{(ij)}$	$e^{(ij)}$	$r_{11}^{(ij)}$	$r_{12}^{(ij)}$	$r_{22}^{(ij)}$	$\hat{Y}^{(ij)}$	$A_1^{(ij)}$	$A_2^{(ij)}$	$m^{(ij)}$	$b^{(ij)}$	$c_{11}^{(ij)}$	$c_{12}^{(ij)}$	$C_{22}^{(ij)}$	$L^{(ij)}$	$\rho^{(ij)}$	$\rho_t^{(j)}$

FIGURE H.1

FRM/MSM

```
0004 DIM E0(10)
0005 DIM X(200,24),X1(200,24),X2(200,24)
0006 DIM X3(200,24)
1000 FILES DR5
1005 DIM I1(200,3)
1010DIM I(16,29),A1(200,5)
1020REM **** N5=NUMBER OF DATA POINTS *************
1022 DIM V9(10),L0(11,17),P0(12,17)
1030 N5=100
```

$\longrightarrow$ number of observations to be processed.

```
1031 V9(1)=.0004
1032 FOR E0=2 TO 10
1033 V9(E0)=V9(E0-1)*1.5
1035 NEXT E0
```

V9(EΦ) for EΦ = 1,2,....10 corresponds to $V_t^{(k)}$ for $k = 1, 2, \cdots 10$ in a V range.

```
1040REM **********   DATA      ********************
1042 MAT READ #1,I1
```

$\longrightarrow$ data stored in matrix I1.

```
1080FOR N=1 TO N5
1085  B0=I1(N,2)
1090 A1(N,1)=EXP(B0)
1095NEXT N
```

```
1110REM ******* CONSTRUCT FIRST 13 COLUMNS OF I  ******
1120FOR I=1 TO 16
1140 I(I,1)=1
1150NEXT I
1160I1=0
1170FOR K=1 TO 4
1180FOR J=1 TO 4
1190I1=I1+1
1200I(I1,2)=J
1210I(I1,3)=K
1220NEXT J
1230NEXT K
```

sets up columns 1, 2, 3 of matrix I

```
1380 MAT L0=ZER
1382 MAT P0=ZER
1400 FOR E0=1 TO 10
1415 P0(E0,17)=1/I0
1490 NEXT E0
1508REM ************* N9=NUMBER OF INITIAL CONDITION SETS *************
1510  N9=1
```

PΦ(EΦ,17) for EΦ = 1, 2, ... 10 corresponds to the prior $p_t^{(k)}$ for $t = 0$ and $k = 1, 2, \cdots 10$

```
1520 K9=0
1522 R2=1
1530 R3=.01
1531 V9=.0025
1533 K2=1.1*(1/R3)
1535 P1=.9
1536 R1=31.3333333
1537 R4=1
1538 K9=K9+1
```

$R2 = 1/\lambda_3$

$R3 = 1/\lambda_4$

$V9 = V_{E,N}$

$K2 = \lambda_2$

$P1 = \pi^{(1)}$

$R1 = \pi^{(2)}/\pi^{(3)}$

$R4 = \pi^{(3)}/\pi^{(4)}$

defines the values of the parameters in the SSP.

```
1540REM **********  INITIAL M,C *********************************
1558FOR I2=1 TO 16
1560I(I2, 9)=5.9
1570 I(I2,10)=.010
1580I(I2,11)=.01
1590I(I2,12)=0
1600I(I2,13)=.0001
1610 NEXT I2
```

sets up columns 9 to 13 of matrix I prior to any observations

FRM/MSM cont.

```
3040W6=V9/R2
```

$W6 = \lambda_2 V_{\varepsilon,N}$

```
3050W7=V9/R3
```

$W7 = \lambda_3 V_{\varepsilon,N}$

```
3070 P2=(R1*R4*(1-P1))/(1+R4+R1*R4)
```

$P2 = \pi^{(2)}$

```
3080 P9=P2/R1
```

$P9 = \pi^{(3)}$

```
3090 P4=P9/R4
```

$P4 = \pi^{(4)}$

```
3100REM ******** CONSTRUCT V,W,GREEK P, P ****************
3110FOR I3=1 TO 4
3200I(I3,7)=I(1+4*(I3-1),8)=P1
3210I(4+I3,7)=I(2+4*(I3-1),8)=P2
3220I(8+I3,7)=I(3+4*(I3-1),8)=P9
3230I(12+I3,7)=I(4+4*(I3-1),8)=P4
```

sets up columns 7 and 8 of matrix I.

```
3240NEXT I3
4000REM ********************************************************
4010REM        OBSERVATIONS   LOOP
4020REM ********************************************************
4023 S0=Q0=B4=B5=0
4025 FOR N1=1 TO N5
4030 N=N1
4032 REM ************** SYSTEM FORECAST *******************
4034 GOSUB 9500
4040REM ***** KALMAN UPDATING ALL TRANSITIONS     *********
4042 Y0=LOG(A1(N,1))
```

$Y0 = y_t$ (current observation at time t).

```
4050GOSUB 9000
4060REM ***** COLLAPSING PROCESS   ************************
4070GOSUB 8000
4072 REM ****  CYCLE INFO GIVEN CURRENT POINT ************
4074 GOSUB 9650
4100 NEXT N1
5000 GOTO 9999
8000REM ****************************************************
8010REM        COLLAPSING  PROCESS
```

Collapsing subroutine.

```
8020REM ****************************************************
8030FOR K=1 TO 4
8035 M1=M2=C1=C2=C3=0
8040FOR J=1 TO 4
8050I1=4*(K-1)+J
8060IF I(I1,1)<>1 THEN 8135
8070I=I1
8080REM ********   COLLAPSE  M  ****************************
8090M3=I(I,22)
8100M4=I(I,23)
8110P0=I(I,28)/I(I,29)
8120M1=P0*M3+M1
8130M2=P0*M4+M2
8135 NEXT J
```

M1 and M2 at this point correspond to $m_t^{(j)}$ and $b_t^{(j)}$ as given in the set of equations (3.3.12)

```
8140I(K,9)=M1
8150I(K,10)=M2
8160REM **********   COLLAPSE  C   *************************
8162 FOR J=1 TO 4
8164 I1=4*(K-1)+J
8166 IF I(I1,1)<>1 THEN 8225
8168 I=I1
8170C4=I(I,24)
8180C5=I(I,25)
8190C6=I(I,26)
```

```
8192 M3=I(I,22)
8193 M4=I(I,23)
8195 P0=I(I,28)/I(I,29)
8200C1=P0*(C4+(M3-M1)**2)+C1
8210C2=P0*(C5+(M3-M1)*(M4-M2))+C2
8220 C3=P0*(C6+(M4-M2)**2)+C3
8225 NEXT J
```

Just after line 8225, C1, C2, C3 correspond to $c^{(j)}_{11,t}$, $c^{(j)}_{12,t}$, $c^{(j)}_{22,t}$ as given in the set of equations (3.3.12).

```
8230I(K,11)=C1
8240I(K,12)=C2
8250I(K,13)=C3
8270NEXT K
8280REM ************
8290I(1,8)=I(1,29)
8300I(2,8)=I(5,29)
8310I(3,8)=I(9,29)
8320I(4,8)=I(13,29)
8330REM  ***
8340REM ***
8350REM *****************************************************
8360FOR K=2 TO 4
8370FOR J=1 TO 4
8380I=4*(K-1)+J
8390IF I(I,1)<>1 THEN 8460
8400I(I,8)=I(J,8)
8410I(I,9)=I(J,9)
8420I(I,10)=I(J,10)
8430I(I,11)=I(J,11)
8440I(I,12)=I(J,12)
8450I(I,13)=I(J,13)
8460NEXT J
8470NEXT K
8480RETURN
```

The posterior (to y_t) information $p^{(j)}_t, m^{(j)}_t$, $b^{(j)}_t$, $c^{(j)}_{11,t}$, $c^{(j)}_{12,t}$, $c^{(j)}_{22,t}$ have now been stored in matrix I between rows 1 and 4 and columns 8 and 13. This set of information can now be considered as our prior (to y_{t+1}) information:

$p^{(i)}_{t-1}$, $m^{(i)}_{t-1}$, $b^{(i)}_{t-1}$, $c^{(i)}_{11,t-1}$, $c^{(i)}_{12,t-1}$, $c^{(i)}_{22,t-1}$

as shown in matrix I.

```
9000REM *****************************************************
9010REM         KALMAN UPDATING
9085 GOSUB 9800
9090L0=0
9095FOR I1=1 TO 16
9100IF I(I1,1)<>1 THEN  9205
9105I=I1
9110I(I,14)=I(I,9)+I(I,10)
9115I(I,15)=Y0-I(I,14)
9120E1=I(I,15)
9125I(I,16)=I(I,11)+2*I(I,12)+I(I,13)+I(I,5)
9130I(I,17)=             I(I,12)+I(I,13)+I(I,6)
9135I(I,18)=                      I(I,13)+I(I,6)
9140I(I,19)=I(I,16)+I(I,4)
9145Y4=1/I(I,19)
9150I(I,20)=Y4*I(I,16)
9155A1=I(I,20)
9160I(I,21)=Y4*I(I,17)
9165A2=I(I,21)
9170I(I,22)=I(I,14)+E1*A1
9175I(I,23)=I(I,10)+E1*A2
9180I(I,24)=I(I,16)*(1-A1)
9185I(I,25)=I(I,17)*(1-A1)
```

This is the Kalman updating subroutine which uses the set of equations (3.3.7) to derive $m^{(ij)}$, $b^{(ij)}$, $c^{(ij)}_{11}$ e.t.c. These are stored in columns 14 to 26 of matrix I. Note that these calculations are performed 16 times, corresponding to the 16 $i \to j$ possible state transitions

FRM/MSM cont.

```
9190I(I,26)=I(I,18)-I(I,19)*A2*A2
9195I(I,27)=SQR(Y4)*EXP(-0.5*E1*E1*Y4)*I(I,7)*I(I,8)
9200L0=L0+I(I,27)
9205NEXT I1
9210 Q0=Q0+LOG(L0)
9215FOR I1=1 TO 16
9220IF I(I1,1)<>1 THEN 9235
9225I=I1
```

→ I(I,27) corresponds to the likelihood product: $L^{(ij)} \cdot \rho^{(i)}_{t-1} \cdot \pi^{(j)}$ as given by equation (3.3.9)

```
9230I(I,28)=I(I,27)/L0
9235NEXT I1
9240FOR I3=1 TO 4
```

→ I(I,28) corresponds to $p^{(ij)}$ as used in (3.3.12)

```
9245L0=0
9250I4=4*(I3-1)
9255FOR I5=1 TO 4
9260IF I(I4+I5,1)<>1 THEN 9270
9265L0=L0+I(I4+I5,28)
9270NEXT I5
9275I(I4+1,29)=I(I4+2,29)=I(I4+3,29)=I(I4+4,29)=L0
9280NEXT I3
9285RETURN
```

The 29th column of I is constructed as shown in figure H.1

```
9500REM *********  SYSTEM FORECASTING  **********************
9505S8=S9=0
9510FOR I=1 TO 4
9515S8=S8+I(I,8)*I(I,9)
9520S9=S9+I(I,8)*I(I,10)
9525NEXT I
9530  IF N<21 THEN 9600
9540  S0=S0+(LOG(A1(N,1))-S6-S7)**2
9600 RETURN
```

Just after line 9525, S8 and S9 correspond to the system level and system growth as defined in section 4.2. Hence their sum S8+S9 is the system forecast. (Note that S8, S9 are equal to S6, S7 below

```
9650 REM ********  CYCLE INFO GIVEN CURRENT POINT  ******
9655S6=S7=0
9660FOR I=1 TO 4
9665S6=S6+I(I,8)*I(I,9)
9670S7=S7+I(I,8)*I(I,10)
9675 NEXT I
9699 RETURN
```

This also calculates m_t and b_t as defined in section 4.2.
ie. S6 = m_t = system level & S7 = b_t = system growth

```
9700 REM ************************************************
9702 W6=V9/R2
9703 W7=V9/R3
9704 FOR I3=1 TO  4
9705I(I3,4)=I(8+I3,4)=I(12+I3,4)=V9
9710I(4+I3,4)=K2*V9
9715I(I3,5)=I(4+I3,5)=0
9720I(12+I3,5)=W7
9725I(8+I3,5)=W6
9730I(I3,6)=I(4+I3,6)=I(12+I3,6)=0
9735I(8+I3,6)=W6
9737 NEXT I3
9740  RETURN
```

Subroutine to set up columns 4,5,6 of matrix I using $\pi^{(j)}$, λ_2, λ_3, λ_4 and $V_{\varepsilon,N}$ to construct $V_\varepsilon^{(j)}$, $V_\gamma^{(j)}$ and $V_\beta^{(j)}$ as shown in table 5.1 in section 5.2.

```
9800REM ************************************************************
9801 MAT L0=ZER
9803 :###
9806 E3=0
9809FOR E0=1 TO 10
9812V9=V9(E0)
```

This is the on line noise variance estimation subroutine.

FRM/MSM cont.

```
9815GOSUB 9700
9818E2=0
9827 FOR I1=1 TO 16
9830 I=I1
9833I(I,14)=I(I,9)+I(I,10)
9836I(I,15)=Y0-I(I,14)
9839E1=I(I,15)
9842I(I,16)=I(I,11)+2*I(I,12)+I(I,13)+I(I,5)
9845I(I,19)=I(I,16)+I(I,4)
9848Y4=1/I(I,19)
9851:     ###.###
9854I(I,27)=SQR(Y4)*EXP(-0.5*E1*E1*Y4)*I(I,7)*I(I,8)
9857 P0(11,I1)=P0(11,17)=P0(12,I1)=0
9860L0(E0,I)=I(I,27)
9863 L0(E0,17)=L0(E0,17)+L0(E0,I)
9869 NEXT I1
9875 NEXT E0
9882FOR E0=1 TO 10
9884L0(E0,17)=(L0(E0,17))*P0(E0,17)
9886P0(11,17)=P0(11,17)+L0(E0,17)
9888NEXT E0
9890V9=0
9892FOR E0=1 TO 10
9894P0(E0,17)=L0(E0,17)/P0(11,17)
9896V9=V9+V9(E0)*P0(E0,17)
9898 NEXT E0
9985GOSUB 9700
9998 RETURN
9999 END
```

The Kalman updating is performed 16 × 10 times corresponding to the 16 $i \to j$ transitions and the 10 $V_\varepsilon^{(k)}$ values in the V range used here.

The likelihoods associated with these 160 "models" are stored in matrix LØ in line 9860.

Finally, V9 just after line 9898 corresponds to the new variance estimate $\hat{V}_{\varepsilon,t}$ which is then used by the Kalman updating routine, in lines 9000 – 9285.

FRM / SSM

```
0100 FILES DR5CV
0200 DIM Y(200,3),Y3(20),V(20),L(20)
0400 DIM P0(20)
0500MAT READ #1,Y
0555 J0=10
0750 FOR U0=1 TO 1
0752IF U0=1 THEN U6=0
0800 R0=25
0810 R9=25
1103 FOR V1=2 TO 2
1107IF V1=2 THEN V(1)=.0004
1110 FOR J=2 TO J0
1115 V(J)=V(J-1)*1.5
1120 P0(J)=1/J0
1130NEXT  J
1140 P0(1)=1/J0
1200 M=6
1300 C=.25
1400 S1=S2=S3=S4=S5=0
1500FOR I=1 TO 200
1600Y0=Y(I,2)
1700Y1=M
1800E1=Y0-Y1
1900 IF I<21 THEN 2200
2000 S1=S1+E1**2
2200L0=0
2300FOR J=1 TO J0
2400Y3(J)=C+V(J)/R9+V(J)
2500L(J)=SQR(1/Y3(J))*EXP(-.5*E1**2/Y3(J))
2700 L0=L0+L(J)
2800NEXT J
2900 P0=0
3000FOR J=1 TO J0
3100 P0(J)=P0(J)*L(J)/L0
3200 P0=P0+P0(J)
3300 NEXT J
3400V=0
3500 FOR J=1 TO J0
3600 P0(J)=P0(J)/P0
3700 V=V+V(J)*P0(J)
3705 P0(J)=U6+P0(J)*(1-J0*U6)
3800 NEXT J
3810 R=C+V/R9
3850 Y3=R+V
4000A=R/Y3
4100M=M+A*E1
4200C=R*(1-A)
4400NEXT I
4600  NEXT V1
5000 NEXT U0
9999 END
```

→ data stored in matrix Y.

→ U6 corresponds to p_L (lower prob. limit)

→ R9 corresponds to the nominated ratio $r_{\mu,N}$

V(J) for J=1,2,....10 corresponds to $V_L^{(k)}$ for k=1,2,....10. Similarly PO(J) corresponds to $p_t^{(k)}$ for t=0 and k=1,2,....10

→ YO = current observation at time t, y_t.

→ forecast error $e_t = y_t - \hat{y}_t$

On line variance estimation routine Y3(J) and L(J) correspond to $\hat{Y}_t^{(k)}$ and $L(y_t|\hat{Y}_t^{(k)}, D_{t-1})$ as given by equations (7.2.1.6) and (7.2.1.7) respectively.

Just after line 3800, V corresponds to our best on line estimate $\hat{V}_{\varepsilon,t}$ as given by equation (7.2.1.12).

PO(J) for J=1,2...10 corresponds to posterior probabilities $p_t^{(k)}$ as given by equation (7.2.1.11).

→ KALMAN UPDATING

R, Y3, A, M, C correspond to R_t, Y_t, A_t, m_t and C_t respectively, as given by equations (7.2.1.13).

VRM / SSM

```
0100 FILES DR5CV
0200 DIM Y(200,3),Y3(2),V(2),L(2)
0400 DIM P0(2)
0410 DIM P1(2)
0500MAT READ #1,Y
0555 J0=2
0800 FOR R0=1 TO 7
0801 R9=25
0900 V1=1.15+R0*.05
1150 V =.0050
1200 M=6
1300 C=.25
1400 S1=S2=S3=S4=S5=0
1500FOR I=1 TO 200
1512 P1(2)=1/(V1+1)
1514 P1(1)=1-P1(2)
1520 V(1)=V *(1/V1)
1530 V(2)=V *V1
1600Y0=Y(I,2)
1700Y1=M
1800E1=Y0-Y1
1900 IF I<21 THEN 2200
2000 S1=S1+E1**2
2200L0=0
2300FOR J=1 TO J0
2400Y3(J)=C+V(J)/R9+V(J)
2500L(J)=SQR(1/Y3(J))*EXP(-.5*E1**2/Y3(J))
2700 L0=L0+L(J)
2800NEXT J
2900 P0=0
3000FOR J=1 TO J0
3100 P0(J)=P1(J)*L(J)/L0
3200 P0=P0+P0(J)
3300 NEXT J
3400V=0
3500 FOR J=1 TO J0
3600 P0(J)=P0(J)/P0
3700 V=V+V(J)*P0(J)
3800 NEXT J
3810 R=C+V/R9
3850 Y3=R+V
4000A=R/Y3
4100M=M+A*E1
4200C=R*(1-A)
4400NEXT I
5000 NEXT R0
9999 END
```

→ data stored in matrix Y

→ R9 corresponds to the nominated ratio $r_{\mu,N}$

→ V1 corresponds to β as defined in section 7.2.4

construction of the variable V range as shown in (7.2.3.3) and (7.2.3.4). P1(1), P1(2), V(1), V(2) correspond to $p^{(1)}, p^{(2)}$, $V_{\varepsilon,t}^{(1)}$ and $V_{\varepsilon,t}^{(2)}$ respectively.

On line variance estimation routine: Y3(J) and L(J) correspond to $\hat{Y}_t^{(k)}$ and $L(y_t \mid \hat{Y}_t^{(k)}, D_{t-1})$ as given by equations (7.2.1.6) and (7.2.3.7) respectively.

→ P0(J) for J=1,2 corresponds to $p_t^{(k)}$ for k=1,2, as given by (7.2.3.8). Just after line 3800, V is the latest on line estimate, $\hat{V}_{\varepsilon,t}$ as given by (7.2.3.9). This estimate is used in the Kalman updating between lines 3810 and 4200.

C I M

```
0100 FILES BJDATA,BJDMSE
0105 DIM Z(1,5),P9(200,16)
0110 DIM V0(200),R9(200),P8(200,9),X(4,5)
0200 DIM Y(200),Y3(20,20),V(20),R0(20),L(20,20),L0(20,20)
0300 DIM  P0(20),Q0(20),S0(20),S1(20),I1(200,1),I2(179)
0400 DIM P(20,20),M(20,20),C(20,20),E1(20,20)
0405 DIM A(20,20),R(20,20)
0410 I0=15
0420 J0=8
0430 X4=.060
0500MAT READ #1,I1
0502 MAT READ #2,I2
0505 FOR X3=1 TO 1
0510 FOR I=1 TO 200
0520 Y(I)=I1(I,X3)
0530 NEXT I
0600FOR X5=3 TO 3
0605IF X5=1 THEN V0=.5*X4
0610IF X5=2 THEN V0=2*X4
0615IF X5=3 THEN V0=4*X4
0650 X0=0
0700 IF X0=0 THEN 800
0705 X0=1
0707 Z(X3,3)=0
0714 A0=V0/4.768371582
0716 V(1)=A0
0718 PRINT "V(1)=";V(1)
0720 FOR I=2 TO I0
0722 V(I)=1.25*V(I-1)
0724 PRINT "V(";I;")=";V(I)
0726 NEXT I
0740 GOTO 1190
0800 R0(1)=1
0850 R0(2)=2
0900 R0(3)=4
1000 R0(4)=8
1120 R0(5)=16
1122 R0(6)=32
1126 R0(7)=64
1128 R0(8)=128
1140 A0=V0/17
1142 V(1)=A0
1150 PRINT "V(1)=";V(1)
1170 FOR I=2 TO I0
1172 V(I)=1.5*V(I-1)
1174 PRINT "V(";I;")=";V(I)
1176 NEXT I
1190 FOR I=1 TO I0
1195 FOR J=1 TO J0
1200 M(I,J)=0
1300 C(I,J)=.1
1350 P(I,J)=1/(I0*J0)
1370 NEXT J
1380 NEXT I
1500FOR N=1 TO 197
1510 IF N<>25 THEN 1600
1512 IF X0=1 THEN 1600
1514 V0=V0(24)
```

B/J Data read and stored in matrix I1.

Construction of $V_t^{(k)}$ and $r_\mu^{(q)}$ as described in section 7.3

V(I) for I = 1,2 .... 15 corresponds to $V_t^{(k)}$ for k = 1,2,....15

and

R0(i) for i = 1,2 ... 8 corresponds to $r_\mu^{(q)}$ for q = 1,2,... 8 .

Specification of $(\mu_{t-1} | M^{(kq)}, D_{t-1})$ for k = 1,2,...15 and q = 1,2... 8 as given by (7.3.2).

P(I,J) for I = 1,...15 + J = 1,... 8 corresponds to $P_{t-1}^{(kq)}$ as given by (7.3.3).

```
1516 GOTO 705
1600Y0=Y(N)
```

→ latest observation at time t, y_t.

```
2200L0=0
2208 FOR K=1 TO 20
2210 S0(K)=0
2220 S1(K)=0
2222 NEXT K
2228IF N<21 THEN 2250
2230B0=0
2232FOR I=1 TO I0
2234FOR J=1 TO J0
2236B0=B0+P(I,J)*M(I,J)
2238NEXT J
2240NEXT I
2242Z(X3,3)=Z(X3,3)+(Y0-B0)**2
2250 FOR J=1 TO J0
2260 FOR I=1 TO I0
2280 E1(I,J)=Y0-M(I,J)
2400R(I,J)=C(I,J)+V(I)/R0(J)
2402Y3(I,J)=R(I,J)+V(I)
2500L(I,J)=SQR(1/Y3(I,J))*EXP(-.5*E1(I,J)**2/Y3(I,J))
2600  L0(I,J)=L(I,J)*P(I,J)
2700 L0=L0+L0(I,J)
2702 S1(I)=S1(I)+L0(I,J)
2703 S0(J)=S0(J)+L0(I,J)
2715A(I,J)=R(I,J)/Y3(I,J)
2720M(I,J)=M(I,J)+A(I,J)*E1(I,J)
2725C(I,J)=R(I,J)*(1-A(I,J))
2750 NEXT I
2800NEXT J
```

Kalman updating calculations in each of the $M^{(kq)}$ models. LO(I,J) corresponds to the likelihoods $L(y_t \mid \hat{Y}_t^{(kq)}, D_{t-1})$ as given by (7.3.8).

```
3000V=0
3005 S5=0
3010R0=0
3020FOR I=1 TO I0
3030FOR J=1 TO J0
3040P(I,J)=L0(I,J)/L0
3050 S5=S5+P(I,J)*M(I,J)
3080NEXT J
3082NEXT I
```

→ P(I,J) at this point corresponds to $p_t^{(kq)}$ as given by (7.3.9).

```
3145 FOR I=1 TO I0
3148 V=V+V(I)*(S1(I)/L0)
3150 P9(N,I)=S1(I)/L0
3228 NEXT I
3230 P9(N,16)=V
3268 V0(N)=V
```

→ V at this point corresponds to $\hat{V}_{\varepsilon,t}$ as given by (7.3.10).

```
3348 FOR J=1 TO J0
3388 R0=R0+R0(J)*(S0(J)/L0)
3390 P8(N,J)=S0(J)/L0
3468 NEXT J
3470 P8(N,9)=R0
3478 R9(N)=R0
3588 NEXT N
3800 NEXT X5
3810 NEXT X3
9999 END
```

→ RO at this point corresponds to $\hat{r}_{\mu,t}$ as given by (7.3.11).

C V M

```
0100 FILES BJDATA
0111 DIM A9(100)
0112 DIM M7(100)
0200 DIM R9(100),P9(100),A(100),V(100),M(100),C(100),S0(100),Y1(100)
0201 DIM E1(100),Y3(100),S1(100),S2(100),L(100)
0210 DIM Y(197,1),E9(100)
0211 DIM X1(200,20),A0(100),Z(100,2)
0212 DIM R(100)
0213 DIM P8(100)
0300 N0=8
0310 U0=0
0312 D7=0
0500MAT READ #1,Y
```

→ data stored in matrix Y.

```
0509 M5=19
0510 V5=.046
0511 PRINT "INITIAL M,V,D7,";M5,V5,D7
0555 FOR K0=1 TO 1
0600 P9(1)=1/N0
0605 R9(1)=1
0606 R0=P9(1)*R9(1)
0610 FOR K=2 TO N0
0615 P9(K)=1/N0
0620 R9(K)=R9(K-1)*2
0621 R0=R0+P9(K)*R9(K)
```

Construction of $r_\mu^{(q)}$ and $p_t^{(q)}$ for $q = 1, 2 \cdots Q$ as described in section 7.4.2. $Q = 8$ here.

P9(K) and R9(K) correspond to $p_t^{(q)}$ and $r_\mu^{(q)}$ respectively.

```
0622 NEXT K
0623 A5=(-1+SQR(1+4*R0))/(2*R0)
0625 FOR K=1 TO N0
0626 A9(K)=(-1+SQR(1+4*R9(K)))/(2*R9(K))
0628 M7(K)=V5/(1-A9(K))
0630 S1(K)=0
1200 M(K)=M5
1250 V(K)=V5
1300 C(K)=100*V(K)
1310 S0(K)=V(K)
1312 C0=C0+P9(K)*C(K)
1313 M3=M3+P9(K)*M(K)
1314 V0=V0+P9(K)*V(K)
1315 NEXT K
1400 S1=S2=S3=S4=S5=0
1500FOR I=1 TO 197
1511:###  ##.##  ##.##   ##.##   ##.##
1512:  ##.###  ##.###   #.###   ###.##
```

Initial values are specified for A9(K), S1(K), M(K), V(K) & C(K) corresponding to A_q, $(MSE)_{t-1}^{(q)}$, $m_{t-1}^{(q)}$, $\hat{V}_{\varepsilon,t-1}^{(q)}$ and $C_{t-1}^{(q)}$ as given by (7.4.3.7), (7.4.3.6), (7.4.3.1) and (7.4.3.3) respectively.

```
1515 Y0=17+Y(I,K0)
```

→ Y0 is the latest observation, y_t

```
1520E0=Y0-M0
1525Y9=C0+V0*(1/R0+1)
1530IF I<11 THEN 1550
1535S4=S4+E0**2
1540S5=S5+LOG(SQR(1/Y9)*EXP(-.5*E0**2/Y9))
1550 L0=0
1560 FOR K=1 TO N0
1700 Y1(K)=M(K)
```

→ Kalman updating:

```
1800 E1(K)=Y0-Y1(K)
1801 IF I=1 THEN E1(K)=1E-10
1803 S1(K)=S1(K)+E1(K)**2
1807 S2(K)=(D7*M7(K)+S1(K))/(D7+I)
```

S2(K) for D7=0 corresponds to $(MSE)_t^{(q)}$ as given by (7.4.3.6)

```
3810 R(K)=C(K)+V(K)/R9(K)
3820 Y3(K)=R(K)+V(K)
3830 A0(K)=R(K)/Y3(K)
3840 V(K)=S2(K)*(1-A9(K))
3851 L(K)=SQR(1/Y3(K))*EXP(-.5*E1(K)**2/Y3(K))
3852 IF L(K)<1E-40 THEN L(K)=1E-40
3853 L0=L0+L(K)
4100 M(K)=M(K)+A0(K)*(Y0-M(K))
4200 C(K)=A0(K)*V(K)
4205NEXT K
4210L1=0
4215FOR K=1 TO N0
4220P9(K)=P9(K)*L(K)/L0
4225L1=L1+P9(K)
4230NEXT K
4232 R0=V0=M0=C0=Y9=0
4235FOR K=1 TO N0
4240P9(K)=P9(K)/L1
4242 IF I>10 THEN 4245
4243 P9(K)=U0+P9(K)*(1-N0*U0)
4245V0=V0+P9(K)*V(K)
4250R0=R0+P9(K)*R9(K)
4251 M0=M0+P9(K)*M(K)
4253 :###     ###.####      ###.##
4255NEXT K
4257 X1(I,(K0-1)*2+1)=V0
4258 X1(I,(K0-1)*2+2)=R0
4300FOR K=1 TO N0
4305C0=C0+P9(K)*(C(K)+(M(K)-M0)**2)
4310NEXT K
4312 Y9=C0+V0*(1/R0+1)
4350 PRINT USING 1511,I,Y0,M3,E0,M0;
4351 PRINT USING 1512,C0,Y9,V0,R0
4360 M3=M0
4400NEXT I
4410 Z(K0,1)=S4/187
4412 Z(K0,2)=S5/187
4444 NEXT K0
6000FOR J=1 TO 1
6005PRINT USING 6010,Z(J,1),Z(J,2)
6010:    #####.#######       #######.#####
6015NEXT J
6020 GOTO 9999
7000 FOR I=1 TO N0
7005 PRINT USING 7010,I,(2)**(I-1),P9(I)
7010 :###      ###.###     ##.####
7020 NEXT I
9999 END
```

R(k), Y3(k), A0(k) correspond to $R_t^{(q)}$, $\hat{Y}_t^{(q)}$ and $A_t^{(q)}$ as given by the set of equations (7.4.3.4)

V(k) corresponds to $V_{\varepsilon,t}^{(q)}$ as given by (7.4.3.5) and L(k) corresponds to $L_t^{(q)}$ as defined in (ii) of section 7.4.3.

M(k) and C(k) in lines 4100 and 4200 correspond to $m_t^{(q)}$, $C_t^{(q)}$ as defined in (iii) of section 7.4.3

Finally, P9(k), V0, R0, M0 correspond to $p_t^{(q)}$, $\hat{V}_{\varepsilon,t}$, $\hat{r}_{p,t}$ and m_t as defined in (ii) section 7.4.3 and the set of equations given in section 7.4.4.

APPENDIX I

FIGURE I.1

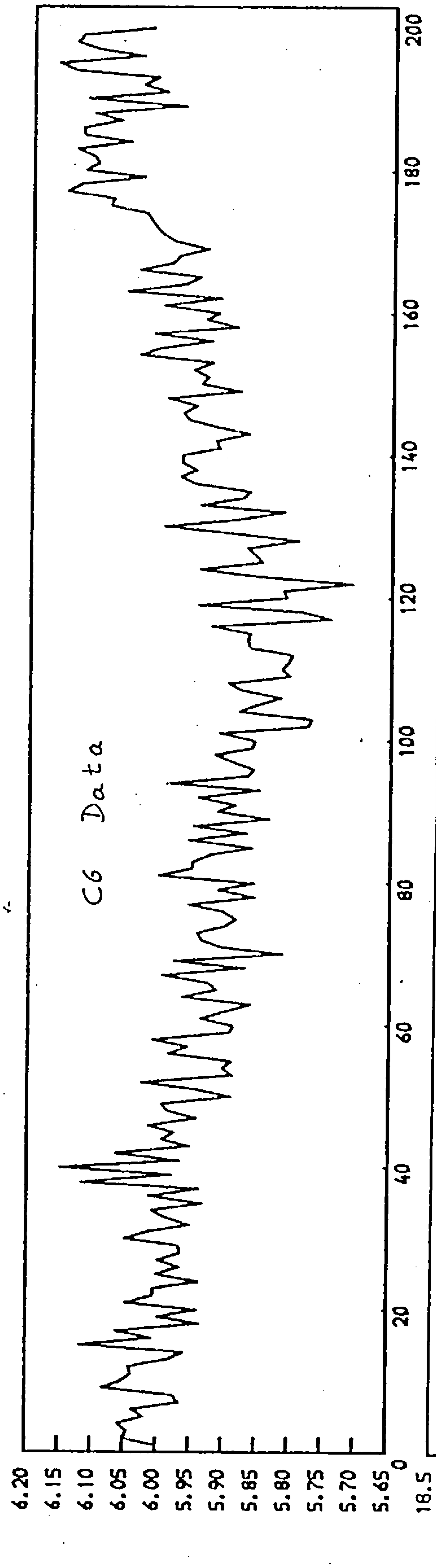


FIGURE I.2

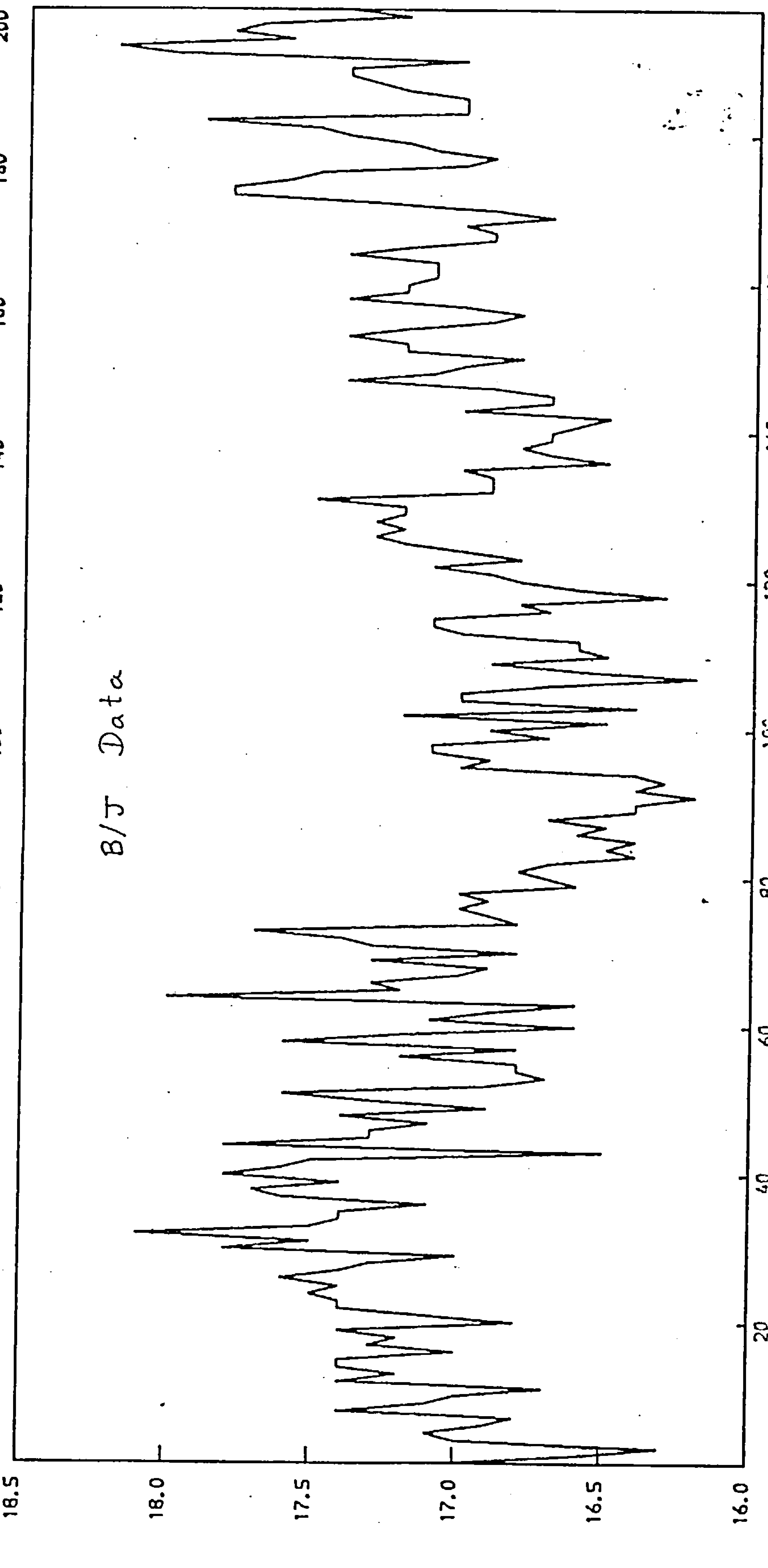
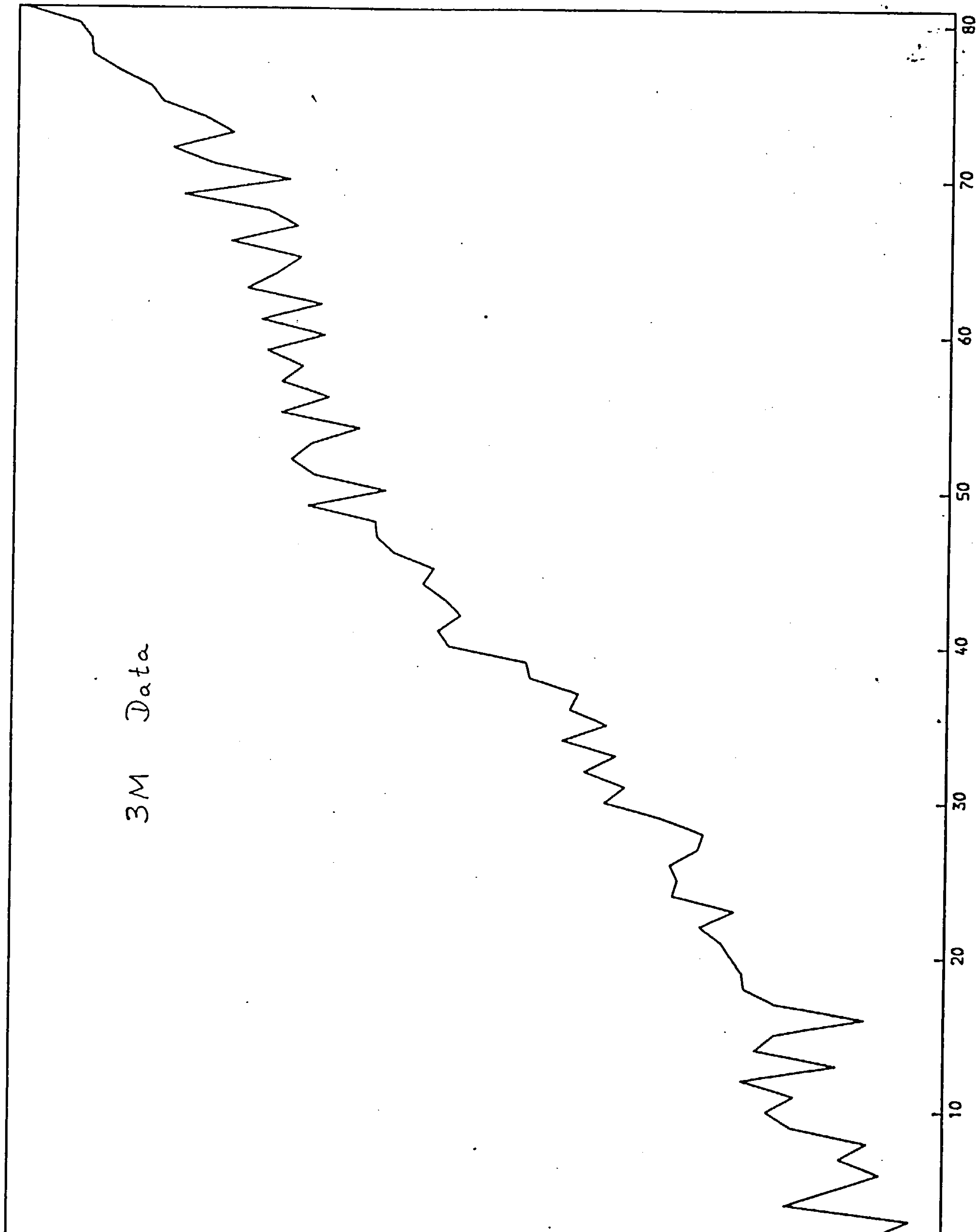


FIGURE I.3



t	$p_t^{(1)}$	$p_t^{(2)}$	$p_t^{(3)}$	$p_t^{(4)}$	$p_t^{(5)}$	$p_t^{(6)}$	$p_t^{(7)}$	$p_t^{(8)}$	$p_t^{(9)}$	$p_t^{(10)}$	$p_t^{(11)}$	$\hat{V}_{\varepsilon,t}$
1	.11	.10	.10	.10	.10	.10	.09	.09	.08	.07	.06	.0102
2	.12	.12	.12	.11	.11	.10	.09	.08	.07	.05	.04	.0083
3	.16	.15	.14	.13	.11	.10	.08	.06	.04	.03	.02	.0058
4	.20	.18	.16	.13	.11	.08	.06	.04	.02	.01	.01	.0040
5	.24	.21	.17	.14	.10	.07	.04	.02	.01	.01		.0030
6	.29	.24	.18	.13	.08	.05	.02	.01				.0023
7	.21	.21	.19	.16	.11	.07	.03	.01	.01			.0027
8	.21	.21	.20	.16	.11	.06	.03	.01				.0025
9	.19	.21	.21	.17	.12	.06	.03	.01				.0025
10	.24	.24	.21	.16	.09	.04	.02					.0021
11	.28	.26	.21	.14	.07	.03	.01					.0018
12	.34	.28	.20	.11	.05	.02						.0016
13	.28	.27	.22	.14	.06	.02						.0017
14	.21	.25	.25	.17	.08	.03	.01					.0019
15	.09	.17	.25	.25	.16	.06	.01					.0025
16	.11	.19	.27	.24	.13	.04	.01					.0023
17	.13	.21	.28	.23	.11	.03	.01					.0022
18	.05	.13	.25	.30	.19	.06	.01					.0027
19	.07	.16	.27	.29	.16	.04	.01					.0025
20	.05	.13	.27	.31	.18	.05	.01					.0026
21	.05	.13	.27	.32	.18	.05	.01					.0026
22	.06	.16	.30	.30	.14	.03						.0024
23	.08	.19	.32	.28	.11	.02						.0022
24	.07	.18	.32	.30	.12	.02						.0022
25	.08	.20	.33	.27	.09	.01						.0021
26	.11	.23	.34	.24	.07	.01						.0019
27	.13	.26	.34	.21	.05	.01						.0018
28	.16	.29	.33	.18	.04							.0017
29	.20	.32	.31	.14	.03							.0016
30	.11	.27	.37	.21	.04							.0018
31	.13	.29	.36	.18	.03							.0017
32	.13	.30	.36	.17	.03							.0017
33	.17	.33	.35	.14	.02							.0016
34	.18	.34	.34	.12	.01							.0015
35	.16	.34	.36	.13	.01							.0016
36	.16	.35	.36	.12	.01							.0016
37	.16	.35	.36	.12	.01							.0015
38	.13	.28	.34	.21	.04							.0018
39	.16	.30	.33	.17	.03							.0017
40	.15	.26	.29	.22	.07	.01						.0018
41	.18	.28	.29	.19	.05							.0017
42	.12	.24	.31	.25	.08	.01						.0019
43	.11	.24	.32	.26	.08							.0019
44	.14	.27	.32	.22	.06							.0018
45	.17	.29	.31	.19	.04							.0017
46	.19	.31	.30	.16	.03							.0016
47	.17	.31	.32	.17	.03							.0017
48	.21	.34	.30	.14	.02							.0015
49	.25	.36	.27	.11	.01							.0015
50	.12	.29	.35	.21	.03							.0018

Posterior probabilities $p_t^{(k)}$ and on line variance estimates $\hat{V}_{\varepsilon,t}$ produced by FRM when applied to C6 Data.

APPENDIX J cont.

t						
51	.14	.31	.34	.17	.02	.0017
52	.10	.27	.38	.22	.03	.0018
53	.05	.21	.40	.29	.05	.0020
54	.05	.22	.41	.28	.04	.0020
55	.05	.22	.41	.28	.04	.0019
56	.03	.19	.42	.31	.05	.0020
57	.04	.21	.43	.28	.04	.0019
58	.02	.16	.43	.34	.05	.0021
59	.02	.15	.43	.35	.05	.0021
60	.02	.16	.44	.34	.04	.0021
61	.02	.18	.45	.31	.03	.0020
62	.03	.21	.46	.27	.03	.0019
63	.03	.21	.47	.27	.02	.0019
64	.02	.16	.47	.32	.03	.0020
65	.02	.19	.48	.28	.02	.0019
66	.03	.22	.49	.25	.02	.0019
67	.01	.15	.49	.33	.03	.0020
68	.01	.14	.49	.33	.03	.0020
69	.01	.12	.48	.36	.03	.0021
70		.06	.40	.48	.06	.0023
71		.07	.43	.45	.04	.0022
72		.08	.45	.43	.04	.0022
73		.09	.47	.41	.03	.0022
74	.01	.10	.49	.38	.03	.0021
75	.01	.12	.51	.35	.02	.0021
76	.01	.14	.53	.31	.02	.0020
77	.01	.13	.53	.32	.02	.0020
78	.01	.12	.53	.33	.02	.0020
79	.01	.14	.55	.29	.01	.0020
80	.01	.15	.56	.28	.01	.0020
81		.07	.49	.41	.02	.0021
82		.08	.50	.40	.02	.0021
83		.09	.52	.37	.02	.0021
84		.10	.55	.33	.01	.0020
85		.09	.54	.36	.01	.0021
86		.09	.55	.35	.01	.0021
87		.08	.55	.36	.01	.0021
88		.09	.56	.34	.01	.0020
89		.06	.52	.40	.01	.0021
90		.07	.55	.37	.01	.0021
91		.08	.57	.33	.01	.0020
92		.08	.58	.32	.01	.0020
93		.08	.58	.33	.01	.0020
94		.05	.53	.41	.01	.0021
95		.05	.54	.40	.01	.0021
96		.05	.55	.39	.01	.0021
97		.06	.58	.35	.01	.0021
98		.07	.60	.32	.01	.0020
99		.08	.62	.30		.0020
100		.09	.63	.27		.0020

APPENDIX J cont.

101	.09	.65	.26		.0020
102	.05	.58	.36	.01	.0021
103	.03	.54	.42	.01	.0022
104	.03	.54	.42	.01	.0021
105	.04	.57	.38	.01	.0021
106	.04	.60	.36		.0021
107	.05	.60	.34		.0021
108	.04	.60	.35		.0021
109	.04	.61	.35		.0021
110	.05	.63	.32		.0020
111	.06	.65	.29		.0020
112	.06	.67	.26		.0020
113	.06	.67	.26		.0020
114	.06	.68	.26		.0020
115	.07	.69	.24		.0020
116	.05	.66	.29		.0020
117	.02	.56	.41		.0021
118	.02	.58	.40		.0021
119	.01	.45	.53	.01	.0023
120	.01	.47	.51	.01	.0022
121	.02	.50	.48	.01	.0022
122	.01	.40	.58	.01	.0023
123	.01	.39	.60	.01	.0023
124		.28	.70	.02	.0024
125		.31	.68	.01	.0024
126		.33	.65	.01	.0024
127	.01	.36	.62	.01	.0023
128		.34	.64	.01	.0024
129		.35	.64	.01	.0024
130		.25	.74	.02	.0025
131		.27	.72	.01	.0024
132		.24	.74	.02	.0025
133		.23	.75	.02	.0025
134		.25	.73	.01	.0025
135		.27	.72	.01	.0024
136		.27	.72	.01	.0024
137		.26	.72	.01	.0024
138		.29	.70	.01	.0024
139		.31	.67	.01	.0024
140		.35	.64	.01	.0024
141		.34	.65	.01	.0024
142		.35	.64	.01	.0024
143		.30	.69	.01	.0024
144		.32	.67	.01	.0024
145		.35	.64	.01	.0024
146		.38	.61		.0023
147	.01	.42	.58		.0023
148	.01	.44	.55		.0023
149		.38	.62		.0023
150		.41	.58		.0023

APPENDIX J cont.

151	.01	.44	.55	.0023
152	.01	.48	.51	.0022
153	.01	.51	.48	.0022
154		.44	.55	.0023
155		.46	.53	.0023
156		.46	.54	.0023
157		.47	.52	.0022
158		.39	.60	.0023
159		.42	.58	.0023
160		.43	.56	.0023
161		.43	.56	.0023
162		.44	.55	.0023
163		.34	.65	.0024
164		.38	.62	.0023
165		.40	.59	.0023
166		.38	.62	.0023
167		.41	.58	.0023
168		.45	.55	.0023
169		.45	.55	.0023
170		.49	.51	.0022
171		.52	.48	.0022
172		.55	.45	.0022
173	.01	.58	.42	.0021
174	.01	.61	.39	.0021
175		.58	.42	.0021
176		.59	.41	.0021
177		.49	.50	.0022
178		.50	.50	.0022
179		.48	.52	.0022
180		.50	.50	.0022
181		.54	.46	.0022
182		.57	.42	.0022
183		.61	.39	.0021
184		.56	.44	.0022
185		.59	.40	.0021
186		.63	.36	.0021
187		.61	.39	.0021
188		.64	.36	.0021
189		.51	.49	.0022
190		.54	.46	.0022
191		.45	.54	.0023
192		.47	.52	.0022
193		.49	.51	.0022
194		.41	.58	.0023
195		.35	.65	.0024
196		.35	.65	.0024
197		.37	.63	.0023
198		.37	.63	.0023
199		.40	.60	.0023
200		.34	.66	.0024

APPENDIX K

RELATIVE IMPORTANCE OF $V_{\varepsilon,N}$ and $r_{\mu,N}$ IN THE SSM

Consider a steady state process with true parameters V_ε and r_μ. These values are unknown and therefore the model employs nominated values $V_{\varepsilon,N}$ and $r_{\mu,N}$. Then at time t the likelihood of a SSM operating with $V_{\varepsilon,N}$ and $r_{\mu,N}$ being the correct model for the process is given by the following log-likelihood function L*.

$$L^* = \sum_t \log L_t \tag{K.1}$$

where $$L_t \propto L(y_t \mid D_{t-1}, V_{\varepsilon,N}, r_{\mu,N}) \tag{K.2}$$

and $$(y_t \mid D_{t-1}, V_{\varepsilon,N}, r_{\mu,N}) \sim N(\hat{y}_t, \hat{Y}_t)$$

Given the observation at time t, y_t, it follows that,

$$\left.\begin{aligned} L_t &\propto [\hat{Y}_t]^{-\frac{1}{2}} \exp\{-\tfrac{1}{2} e_t^2 / \hat{Y}_t\} \\ \text{where} \quad e_t &= y_t - \hat{y}_t \end{aligned}\right\} \tag{K.3}$$

From (K.1) and K.2) it follows that,

$$L^* \propto L(D_t \mid V_{\varepsilon,N}, r_{\mu,N})$$

and therefore L* represents the likelihood of $D_t = \{y_1\ y_2\ \ldots .\ y_t\}$ having been generated by a SSM with constant variances $V_{\varepsilon,N}$ and $V_{\mu,N} = V_{\varepsilon,N} / r_{\mu,N}$.

L* can then be viewed as the optimal performance criterion for the SSM since in the limit it is maximised only when both our nominated estimates $V_{\varepsilon,N}$ and $r_{\mu,N}$ are equal to the true parameters V_ε and r_μ respectively. This is in contrast with the traditional MSE criterion of performance which can be minimised by $r_{\mu,N} = r_\mu$ and ignores $V_{\varepsilon,N}$ completely.

Given that L* is the optimal performance criterion for the SSM, we now examine its sensitivity to $V_{\varepsilon,N}$ and $r_{\mu,N}$. Before this can be done however we need to derive an expression for the limiting value of L* in terms of the nominated and true parameters.

When the model is in its limiting form, the limiting value of $\hat{Y}_t$ is $\hat{Y}_N = V_{\varepsilon,N} / (1 - A_N)$ as given in section 2.3 and if this is substituted in (K.3) we get:

$$L_t \propto \left[\frac{V_{\varepsilon,N}}{1-A_N}\right]^{-\frac{1}{2}} \cdot \exp\left\{ -\tfrac{1}{2}\frac{(1-A_N)}{V_{\varepsilon,N}} e_t^2 \right\}$$

Using this together with the expression for the expected value of e_t^2 as given in section 2.3 we get:

$$E(\log L_t) = E^* \propto -\tfrac{1}{2}\left\{\log\left[\frac{V_{\varepsilon,N}}{1-A_N}\right] + \left[\frac{1-A_N}{V_{\varepsilon,N}}\right]\left[\frac{2A_N V_\varepsilon + V_\mu}{A_N(2-A_N)}\right]\right\}$$

Clearly maximization of L* is in the limit equivalent to maximisation of E* and therefore the latter can be used to examine the sensitivity of SSM's performance to the choice of $V_{\varepsilon,N}$ and $r_{\mu,N}$.

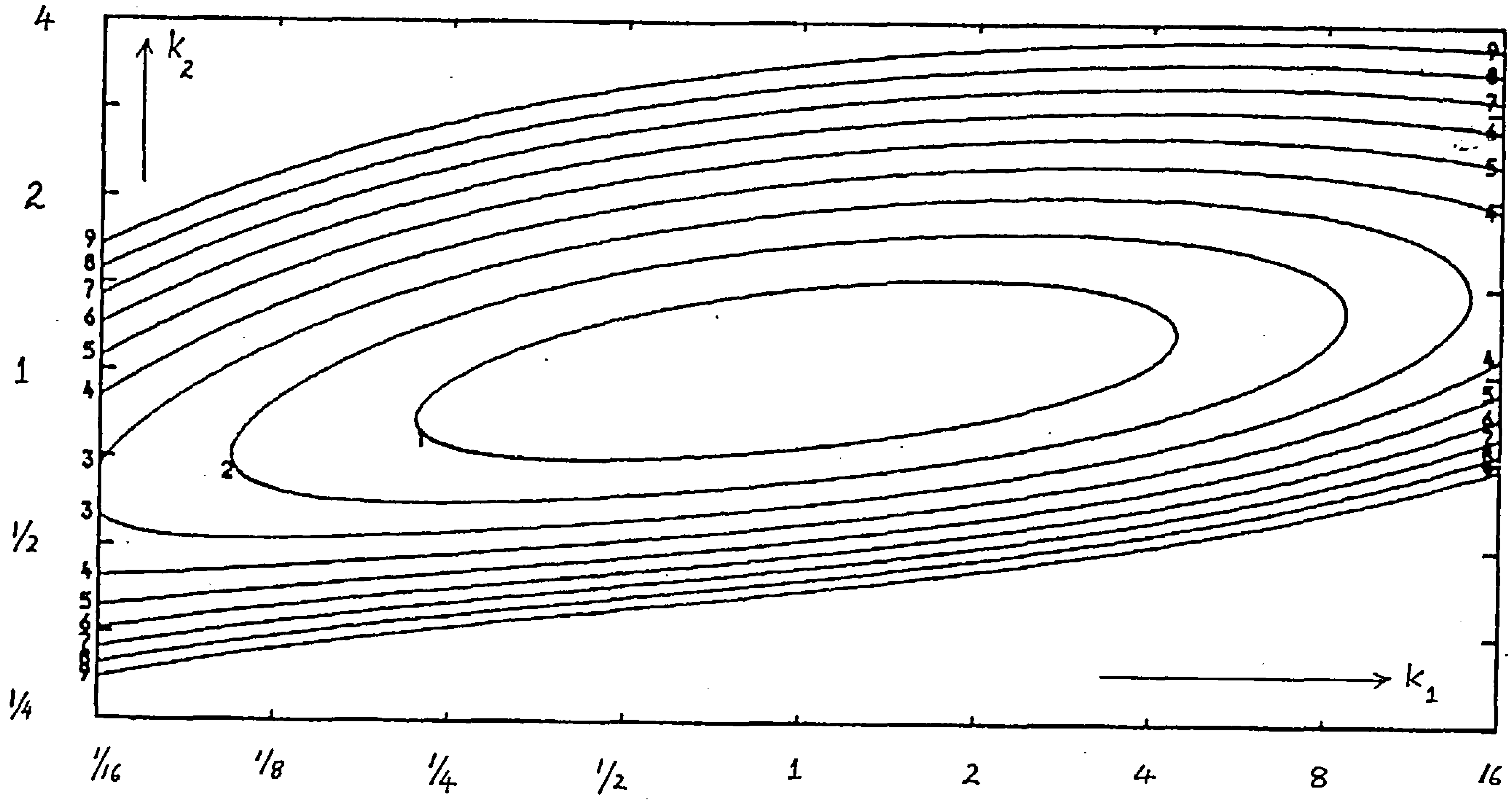
Contours of E* for a typical steady state process with $V_\varepsilon = 100$ and $V_\mu = 5$ (i.e. $r_\mu = 20$ corresponding to an EWMA optimal $\alpha = 0.2$) are given in figure K.1 below. The horizontal and vertical axis in units of k_1 and k_2 represent the factors by which $r_{\mu,N}$ and $V_{\varepsilon,N}$ respectively overestimate or underestimate the true process parameters, i.e.

$$r_{\mu,N} = k_1 \, r_\mu$$

$$V_{\varepsilon,N} = k_2 \, V_\varepsilon$$

When $k_1 = k_2 = 1$ then $r_{\mu,N} = r_\mu = 20$ and $V_{\varepsilon,N} = V_\varepsilon = 100$ thus maximising the expected value of $\log L_t$.

FIGURE K.1 (r=20)



Contours of E* with $V_\varepsilon = 100$, $r_\mu = 20$, $V_{\varepsilon,N} = k_2\, V_\varepsilon$, $r_{\mu,N} = k_1\, r_\mu$

Contour i (i = 1, 2, ... 9) shows i % decrease in E* from the maximum likelihood value of E* when $k_1 = k_2 = 1$.

It is clear from figure K.1 that E* is much more sensitive to $V_{\varepsilon,N}$ than to $r_{\mu,N}$. For example overestimation of r_μ by a factor if 16 ($k_1 = 16$) results in a 3% penalty (Contour line 3) provided k_2 is selected to maximise E* given k_1, while the same penalty is paid for overestimating V_ε by a factor of only 2 ($k_2 = 2$) and selecting k_1 to maximise E* given k_2. Hence the performance of the SSM is relatively insensitive to the choice of $r_{\mu,N}$ and consequently fixing $r_{\mu,N}$ initially will not be critical to SSM's performance provided $V_{\varepsilon,N}$ can be chosen to maximise E* given $r_{\mu,N}$. This is exactly what FRM and VRM are designed to do.

Finally fitures K.2 and K.3 show contours of E* corresponding to $r_\mu=40$(EWMA $\alpha = .15$) and $r_\mu = 10$ (EWMA $\alpha = .27$) respectively, with $V_\varepsilon = 100$ in both cases. It can be seen that the choice of $r_{\mu,N}$ becomes more significant as r_μ decreases and therefore FRM and VRM are likely to be more accurate for large r_μ (say $r_\mu > 10$). When r_μ is very small ($r_\mu \ll 10$) then it may be better to estimate both V_ε and r_μ using the on line procedure described in section 7.3.

FIGURE K.2 (r=40)

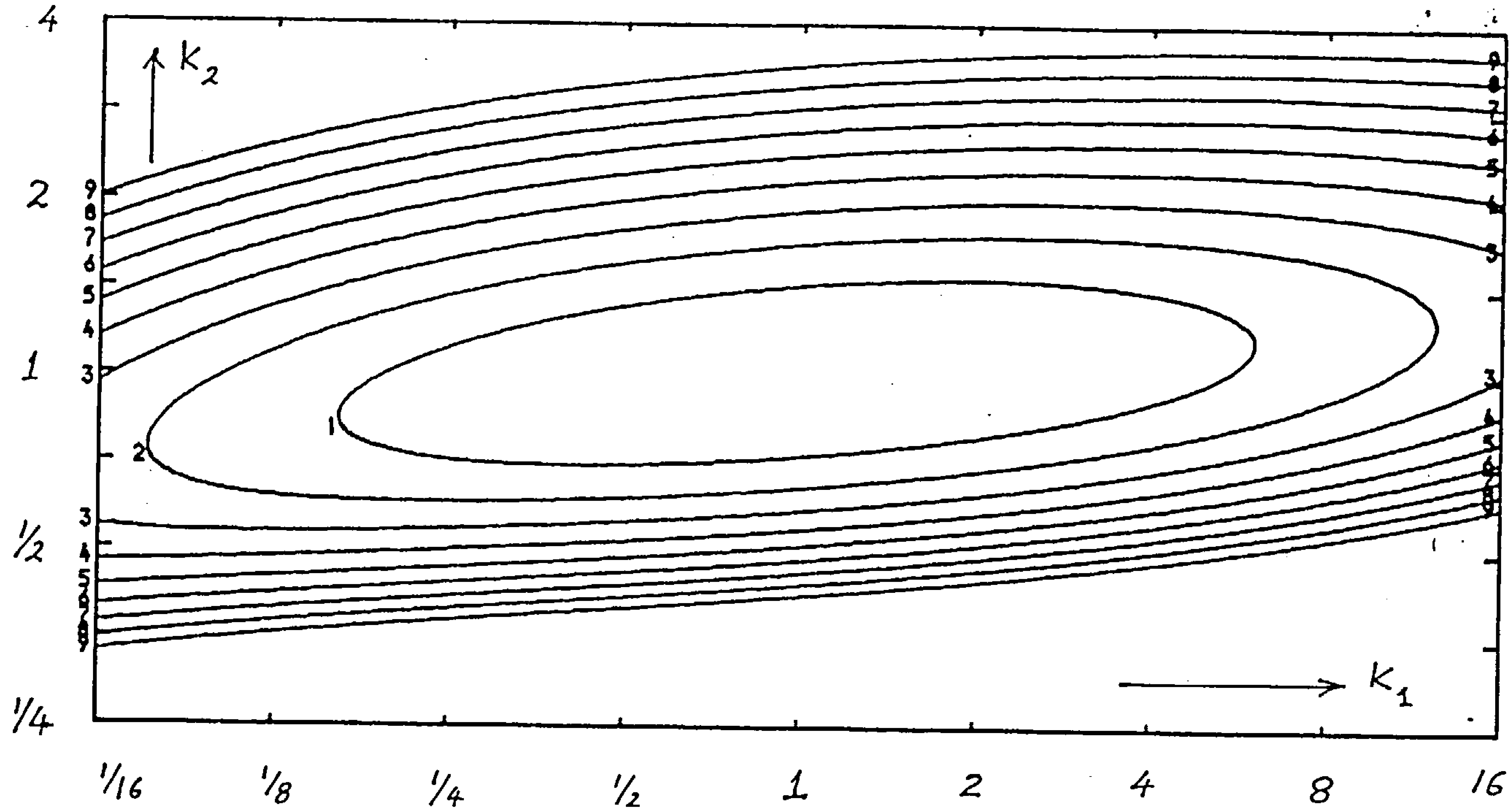
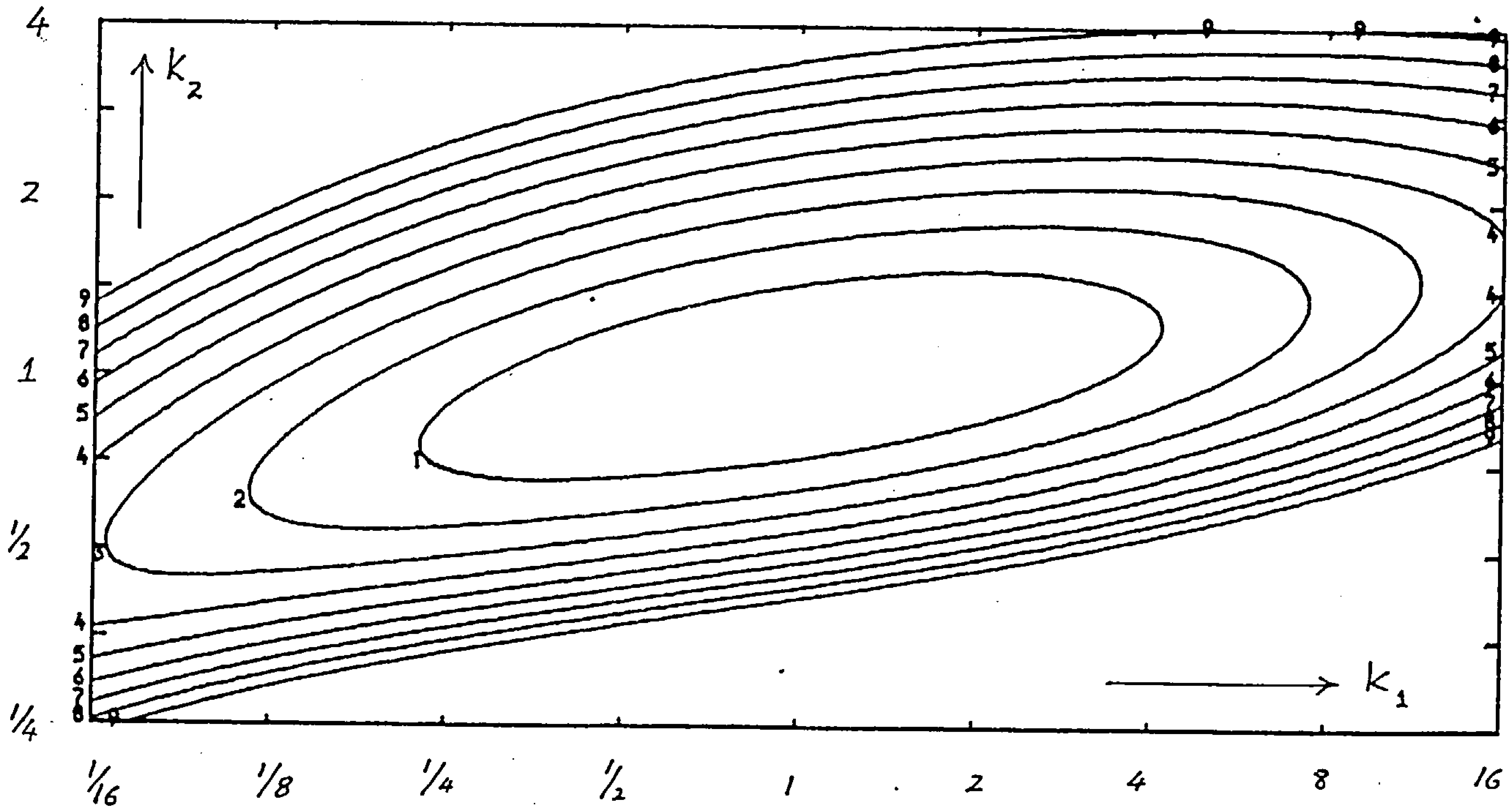


FIGURE K.3 (r=10)



APPENDIX L

Table L.1

t	y_t	$\hat{y}_t$	e_t	m_t	C_t	$\hat{Y}_t$	$\hat{V}_{\varepsilon,t}$	$\hat{r}_{\mu,t}$
1	17.00	19.00	-2.00	17.02	.000	.000	.000	31.91
2	16.60	17.02	-.42	16.79	.033	.097	.062	31.91
3	16.30	16.79	-.49	16.58	.043	.143	.098	31.60
4	16.10	16.58	-.48	16.39	.048	.162	.110	27.67
5	17.10	16.39	.71	16.65	.049	.222	.169	42.11
6	16.90	16.65	.25	16.72	.041	.194	.149	41.14
7	16.80	16.72	.08	16.73	.034	.165	.128	40.41
8	17.40	16.73	.67	16.92	.053	.207	.150	36.03
9	17.10	16.92	.18	16.96	.045	.184	.135	32.55
10	17.00	16.96	.04	16.96	.036	.161	.121	30.90
11	16.70	16.96	-.26	16.87	.029	.150	.117	33.80
12	17.40	16.87	.53	17.02	.038	.166	.124	31.86
13	17.20	17.02	.18	17.07	.036	.154	.114	28.35
14	17.40	17.07	.33	17.17	.041	.152	.107	22.73
15	17.40	17.17	.23	17.25	.039	.143	.098	17.83
16	17.00	17.25	-.25	17.14	.030	.132	.098	20.15
17	17.30	17.14	.16	17.19	.029	.126	.092	17.91
18	17.20	17.19	.01	17.19	.027	.118	.086	16.68
19	17.40	17.19	.21	17.26	.028	.115	.081	14.53
20	16.80	17.26	-.46	17.10	.026	.119	.088	19.12
21	17.10	17.10	.00	17.10	.024	.112	.084	18.53
22	17.40	17.10	.30	17.20	.025	.112	.083	18.15
23	17.40	17.20	.20	17.26	.026	.110	.079	16.36
24	17.50	17.26	.24	17.34	.027	.109	.076	14.02
25	17.40	17.34	.06	17.35	.025	.103	.072	12.70
26	17.60	17.35	.25	17.44	.027	.102	.069	10.65
27	17.40	17.44	-.04	17.42	.023	.096	.066	10.23
28	17.30	17.42	-.12	17.37	.021	.092	.065	10.29
29	17.00	17.37	-.37	17.24	.023	.095	.066	11.34
30	17.80	17.24	.56	17.43	.024	.104	.074	13.20
31	17.50	17.43	.07	17.45	.023	.100	.071	12.35
32	18.10	17.45	.65	17.68	.035	.117	.075	9.91
33	17.50	17.68	-.18	17.60	.026	.108	.075	10.79
34	17.40	17.60	-.20	17.53	.023	.104	.074	11.40
35	17.40	17.53	-.13	17.48	.022	.101	.073	11.43
36	17.10	17.48	-.38	17.36	.025	.104	.073	11.80
37	17.60	17.36	.24	17.44	.022	.101	.074	12.46
38	17.70	17.44	.26	17.52	.022	.101	.073	12.22
39	17.40	17.52	-.12	17.48	.021	.098	.072	12.41
40	17.80	17.48	.32	17.58	.022	.100	.071	12.04
41	17.60	17.58	.02	17.58	.021	.097	.069	11.76
42	17.50	17.58	-.08	17.55	.020	.094	.068	11.82
43	16.50	17.55	-1.05	17.24	.038	.130	.087	16.46
44	17.80	17.24	.56	17.42	.022	.123	.096	22.73
45	17.30	17.42	-.12	17.39	.022	.120	.094	22.44
46	17.30	17.39	-.09	17.37	.022	.118	.092	22.05
47	17.10	17.37	-.27	17.31	.023	.118	.091	21.23
48	17.40	17.31	.09	17.33	.021	.115	.089	21.50
49	16.90	17.33	-.43	17.23	.024	.118	.090	20.31
50	17.30	17.23	.07	17.25	.022	.115	.088	20.54

CVM applied to B/J Data

APPENDIX L cont.

Table L.1 cont.

51	17.60	17.25	.35	17.34	.021	.114	.089	22.24
52	16.90	17.34	-.44	17.23	.022	.117	.091	22.34
53	16.70	17.23	-.53	17.10	.027	.123	.092	19.80
54	16.80	17.10	-.30	17.02	.028	.123	.089	17.00
55	16.80	17.02	-.22	16.96	.028	.121	.087	14.90
56	17.20	16.96	.24	17.03	.023	.116	.087	16.37
57	16.80	17.03	-.23	16.97	.024	.116	.086	15.20
58	17.60	16.97	.63	17.14	.023	.120	.092	19.05
59	17.20	17.14	.06	17.15	.022	.117	.090	18.91
60	16.60	17.15	-.55	17.02	.023	.122	.093	19.37
61	17.10	17.02	.08	17.04	.022	.119	.092	19.52
62	16.90	17.04	-.14	17.01	.022	.117	.091	18.93
63	16.60	17.01	-.41	16.91	.024	.120	.090	17.56
64	18.00	16.91	1.09	17.16	.024	.136	.108	27.39
65	17.20	17.16	.04	17.17	.023	.133	.107	27.17
66	17.30	17.17	.13	17.19	.023	.132	.105	26.65
67	17.00	17.19	-.19	17.15	.021	.129	.104	27.22
68	16.90	17.15	-.25	17.10	.021	.129	.104	27.32
69	17.30	17.10	.20	17.14	.021	.127	.103	27.57
70	16.80	17.14	-.34	17.07	.021	.127	.103	27.71
71	17.30	17.07	.23	17.12	.020	.126	.102	28.29
72	17.40	17.12	.28	17.18	.021	.126	.102	28.28
73	17.70	17.18	.52	17.29	.024	.130	.103	27.09
74	16.80	17.29	-.49	17.18	.020	.129	.105	30.79
75	16.90	17.18	-.28	17.12	.020	.129	.105	31.03
76	17.00	17.12	-.12	17.10	.020	.127	.104	30.66
77	16.90	17.10	-.20	17.06	.021	.126	.102	29.99
78	17.00	17.06	-.06	17.05	.020	.125	.101	29.57
79	16.60	17.05	-.45	16.96	.023	.127	.101	27.82
80	16.70	16.96	-.26	16.90	.024	.127	.100	25.79
81	16.80	16.90	-.10	16.88	.023	.125	.098	24.73
82	16.70	16.88	-.18	16.84	.023	.124	.097	23.26
83	16.40	16.84	-.44	16.74	.026	.126	.096	20.39
84	16.50	16.74	-.24	16.68	.026	.125	.094	18.44
85	16.40	16.68	-.28	16.61	.026	.124	.092	16.50
86	16.60	16.61	-.01	16.61	.024	.121	.091	16.19
87	16.50	16.61	-.11	16.58	.024	.119	.090	15.53
88	16.70	16.58	.12	16.62	.022	.117	.089	15.80
89	16.40	16.62	-.22	16.56	.023	.117	.088	15.20
90	16.40	16.56	-.16	16.52	.023	.116	.087	14.66
91	16.20	16.52	-.32	16.44	.024	.116	.086	13.78
92	16.40	16.44	-.04	16.43	.023	.114	.085	13.56
93	16.30	16.43	-.13	16.40	.023	.113	.084	13.17
94	16.40	16.40	.00	16.40	.022	.112	.083	13.10
95	17.00	16.40	.60	16.56	.022	.114	.086	14.30
96	16.90	16.56	.34	16.65	.023	.115	.086	14.00
97	17.10	16.65	.45	16.77	.026	.118	.085	13.06
98	17.10	16.77	.33	16.86	.027	.118	.084	11.89
99	16.70	16.86	-.16	16.81	.024	.114	.084	12.44
100	16.90	16.81	.09	16.83	.023	.113	.083	12.18

Table L.1 cont.

101	16.50	16.83	-.33	16.74	.022	.112	.084	12.85
102	17.20	16.74	.46	16.86	.023	.114	.084	12.78
103	16.40	16.86	-.46	16.74	.022	.114	.086	13.82
104	17.00	16.74	.26	16.81	.022	.114	.086	13.86
105	17.00	16.81	.19	16.86	.022	.114	.085	13.64
106	16.70	16.86	-.16	16.81	.022	.112	.085	13.84
107	16.20	16.81	-.61	16.66	.023	.116	.087	14.30
108	16.60	16.66	-.06	16.64	.022	.114	.086	14.14
109	16.90	16.64	.26	16.71	.021	.113	.086	14.60
110	16.50	16.71	-.21	16.66	.022	.113	.086	14.52
111	16.60	16.66	-.06	16.64	.021	.112	.085	14.41
112	16.60	16.64	-.04	16.63	.021	.111	.084	14.31
113	17.00	16.63	.37	16.73	.021	.111	.084	14.71
114	17.10	16.73	.37	16.82	.022	.112	.084	14.43
115	17.10	16.82	.28	16.89	.023	.112	.083	13.80
116	16.70	16.89	-.19	16.84	.021	.111	.083	14.31
117	16.80	16.84	-.04	16.83	.021	.109	.083	14.31
118	16.30	16.83	-.53	16.69	.022	.111	.084	14.85
119	16.60	16.69	-.09	16.67	.021	.110	.083	14.66
120	16.80	16.67	.13	16.71	.021	.109	.083	14.84
121	16.90	16.71	.19	16.76	.020	.109	.083	14.95
122	17.10	16.76	.34	16.84	.021	.109	.083	14.83
123	16.80	16.84	-.04	16.83	.020	.108	.082	14.86
124	17.00	16.83	.17	16.87	.021	.108	.081	14.63
125	17.20	16.87	.33	16.96	.022	.108	.081	14.08
126	17.30	16.96	.34	17.05	.023	.109	.080	13.17
127	17.20	17.05	.15	17.09	.023	.108	.079	12.63
128	17.30	17.09	.21	17.15	.023	.108	.078	11.98
129	17.20	17.15	.05	17.16	.022	.106	.078	11.76
130	17.20	17.16	.04	17.17	.022	.105	.077	11.60
131	17.50	17.17	.33	17.26	.023	.106	.076	11.02
132	16.90	17.26	-.36	17.16	.021	.105	.077	11.84
133	16.90	17.16	-.26	17.09	.021	.104	.077	12.00
134	16.90	17.09	-.19	17.04	.021	.104	.077	11.92
135	17.00	17.04	-.04	17.03	.021	.103	.076	11.88
136	16.50	17.03	-.53	16.88	.023	.106	.077	11.51
137	16.70	16.88	-.18	16.83	.023	.105	.076	11.05
138	16.80	16.83	-.03	16.82	.022	.104	.075	10.94
139	16.70	16.82	-.12	16.79	.022	.103	.074	10.68
140	16.70	16.79	-.09	16.77	.021	.102	.074	10.49
141	16.60	16.77	-.17	16.72	.022	.102	.073	10.21
142	16.50	16.72	-.22	16.65	.022	.102	.072	9.84
143	17.00	16.65	.35	16.76	.021	.101	.073	10.45
144	16.70	16.76	-.06	16.74	.020	.100	.073	10.41
145	16.70	16.74	-.04	16.73	.020	.099	.072	10.36
146	16.90	16.73	.17	16.78	.020	.099	.072	10.43
147	17.40	16.78	.62	16.96	.022	.102	.073	10.39
148	17.10	16.96	.14	17.00	.022	.102	.072	10.08
149	17.00	17.00	.00	17.00	.021	.100	.072	10.05
150	16.80	17.00	-.20	16.94	.020	.099	.072	10.30

Table L.1 cont.

151	17.20	16.94	.26	17.01	.021	.099	.072	10.18
152	17.20	17.01	.19	17.07	.021	.099	.071	9.95
153	17.40	17.07	.33	17.17	.022	.100	.070	9.49
154	17.20	17.17	.03	17.17	.021	.099	.070	9.39
155	16.90	17.17	-.27	17.09	.020	.098	.070	9.77
156	16.80	17.09	-.29	17.01	.020	.098	.070	9.87
157	17.00	17.01	-.01	17.00	.020	.097	.070	9.85
158	17.40	17.00	.40	17.12	.020	.098	.070	10.02
159	17.20	17.12	.08	17.14	.020	.097	.070	9.93
160	17.20	17.14	.06	17.16	.020	.097	.069	9.86
161	17.10	17.16	-.06	17.14	.020	.096	.069	9.89
162	17.10	17.14	-.04	17.13	.020	.095	.069	9.89
163	17.10	17.13	-.03	17.12	.019	.095	.068	9.88
164	17.40	17.12	.28	17.20	.020	.095	.068	9.86
165	17.20	17.20	-.00	17.20	.019	.094	.068	9.84
166	16.90	17.20	-.30	17.11	.019	.094	.068	10.03
167	16.90	17.11	-.21	17.05	.020	.094	.068	9.94
168	17.00	17.05	-.05	17.04	.019	.093	.067	9.88
169	16.70	17.04	-.34	16.94	.020	.094	.067	9.62
170	16.90	16.94	-.04	16.93	.020	.093	.066	9.54
171	17.30	16.93	.37	17.04	.019	.093	.067	9.99
172	17.80	17.04	.76	17.26	.022	.098	.069	9.93
173	17.80	17.26	.54	17.43	.026	.101	.068	8.65
174	17.60	17.43	.17	17.49	.024	.099	.067	8.11
175	17.50	17.49	.01	17.49	.022	.097	.066	8.02
176	17.00	17.49	-.49	17.33	.021	.096	.068	8.91
177	16.90	17.33	-.43	17.20	.021	.098	.069	8.97
178	17.10	17.20	-.10	17.17	.021	.097	.068	8.86
179	17.20	17.17	.03	17.18	.021	.096	.068	8.88
180	17.40	17.18	.22	17.25	.020	.095	.068	9.02
181	17.50	17.25	.25	17.33	.021	.096	.068	8.97
182	17.90	17.33	.57	17.51	.023	.099	.068	8.48
183	17.00	17.51	-.51	17.34	.020	.098	.070	9.67
184	17.00	17.34	-.34	17.24	.021	.098	.070	9.76
185	17.00	17.24	-.24	17.17	.021	.098	.070	9.54
186	17.20	17.17	.03	17.18	.021	.097	.069	9.57
187	17.30	17.18	.12	17.22	.020	.096	.069	9.66
188	17.40	17.22	.18	17.27	.020	.096	.069	9.70
189	17.40	17.27	.13	17.31	.020	.096	.069	9.64
190	17.00	17.31	-.31	17.22	.020	.096	.069	9.85
191	18.00	17.22	.78	17.44	.022	.100	.071	10.33
192	18.20	17.44	.76	17.68	.028	.107	.071	8.77
193	17.60	17.68	-.08	17.65	.023	.102	.071	9.00
194	17.80	17.65	.15	17.70	.023	.101	.070	8.62
195	17.70	17.70	.00	17.69	.022	.100	.070	8.57
196	17.20	17.69	-.49	17.54	.021	.100	.071	9.29
197	17.40	17.54	-.14	17.50	.021	.100	.071	9.25

APPENDIX M

TABLE M.1a

t	REALISATION 1	2	3	4	5	6	7	8	9	10
1	.004	-.003	-.001	.006	.000	.002	-.002	-.004	-.003	-.006
2	.000	.003	-.008	-.015	.009	.012	.002	-.001	-.005	-.001
3	.001	.012	.005	.012	.010	-.006	-.001	.020	.009	-.007
4	-.006	.008	.009	.008	.011	-.015	-.000	.017	.006	.002
5	.010	-.007	.014	-.001	.004	-.020	.004	.008	.000	.003
6	.007	-.003	.013	-.011	.003	-.017	.012	.001	.004	.008
7	.007	-.009	.014	-.008	-.005	-.014	.007	.004	-.006	.010
8	.005	-.008	.007	-.016	-.007	-.007	.002	.009	-.008	.004
9	.006	-.004	.008	-.009	-.001	-.000	.007	.010	-.006	.003
10	.000	-.001	.008	-.006	.001	-.003	.005	.004	-.006	.006
11	-.002	-.004	.005	-.007	.001	.005	.006	.005	-.001	.002
12	-.001	-.002	.006	-.004	.002	-.001	.001	.002	-.005	-.001
13	-.003	-.005	.004	-.005	-.001	-.004	.003	.002	-.001	-.002
14	-.001	-.005	.001	-.001	-.003	-.005	-.001	.003	.001	-.000
15	-.005	-.005	.002	-.005	.001	-.006	-.002	.004	-.003	.001
16	-.007	-.002	.002	-.006	-.000	-.006	-.004	.005	-.002	.003
17	-.009	-.001	.002	-.003	.001	-.005	-.003	.006	-.003	.004
18	-.009	-.000	.001	-.002	-.002	-.006	-.002	.004	.001	.005
19	-.005	-.000	-.001	-.002	-.002	-.006	-.000	.001	.002	.003
20	-.004	-.002	.001	-.001	-.004	-.007	-.002	.004	-.000	.002
21	-.004	-.001	-.000	-.000	-.001	-.005	-.003	.004	.002	.000
22	-.003	.000	.001	.003	.000	-.003	.001	.005	.001	-.000
23	.002	.002	.002	.005	.002	-.001	.000	.008	.004	.002
24	.018	.007	.005	.008	.002	.001	.003	.010	.006	.003
25	.001	.015	.005	.011	.005	.003	.005	.016	.010	.004
26	.011	.029	.009	.012	.007	.002	.005	.014	.010	.006
27	.024	.024	.014	.013	.010	.006	.006	.012	.014	.008
28	.035	.020	.020	.014	.012	.010	.011	.012	.015	.011
29	.033	.026	.021	.015	.014	.018	.017	.010	.020	.013
30	.042	.026	.023	.019	.019	.030	.022	.015	.026	.013
31	.038	.023	.024	.019	.023	.034	.021	.018	.028	.013
32	.040	.022	.024	.022	.022	.042	.021	.021	.026	.014
33	.036	.023	.028	.020	.024	.037	.023	.022	.029	.016
34	.033	.024	.022	.018	.026	.038	.020	.024	.029	.017
35	.035	.025	.023	.021	.024	.031	.022	.024	.027	.016
36	.034	.027	.023	.024	.026	.029	.026	.028	.028	.015
37	.033	.024	.022	.023	.024	.030	.028	.027	.027	.014
38	.032	.022	.022	.025	.029	.026	.026	.024	.029	.018
39	.031	.024	.022	.027	.027	.022	.023	.023	.030	.018
40	.029	.026	.023	.033	.033	.026	.025	.024	.031	.021
41	.029	.026	.021	.032	.028	.027	.023	.023	.031	.020
42	.029	.024	.024	.026	.031	.028	.023	.023	.028	.021
43	.030	.024	.023	.024	.027	.033	.022	.022	.027	.024
44	.030	.028	.023	.021	.027	.031	.020	.025	.023	.021
45	.029	.027	.024	.026	.025	.033	.018	.024	.023	.021
46	.028	.027	.025	.028	.026	.028	.022	.025	.018	.019
47	.028	.028	.025	.029	.024	.031	.020	.027	.025	.020
48	.027	.028	.023	.030	.024	.031	.020	.027	.024	.023
49	.027	.029	.026	.026	.024	.030	.021	.028	.021	.033
50	.029	.029	.025	.025	.022	.030	.020	.031	.021	.023

b_t (estimate of β_t) produced by MSM to a growth change of size $\frac{1}{2}\sigma$ at $t = 21$

i.e. $\beta_t = \begin{cases} 0 & t < 21 \\ 0.025 & t \geqslant 21 \end{cases}$

TABLE M.1b

t	REALISATION 1	2	3	4	5	6	7	8	9	10
1	.004	-.003	-.001	.006	.000	.002	-.002	-.004	-.003	-.006
2	.000	.003	-.008	-.015	.009	.012	.002	-.001	-.005	-.001
3	.001	.012	.005	.012	.010	-.006	-.001	.020	.009	-.007
4	-.006	.008	.009	.008	.011	-.015	-.000	.017	.006	.002
5	.010	-.007	.014	-.001	.004	-.020	.004	.008	.000	.003
6	.007	-.003	.013	-.011	.003	-.017	.012	.001	.004	.008
7	.007	-.009	.014	-.008	-.005	-.014	.007	.004	-.006	.010
8	.005	-.008	.007	-.016	-.007	-.007	.002	.009	-.008	.004
9	.006	-.004	.008	-.009	-.001	-.000	.007	.010	-.006	.003
10	.000	-.001	.008	-.006	.001	-.003	.005	.004	-.006	.006
11	-.002	-.004	.005	-.007	.001	.005	.006	.005	-.001	.002
12	-.001	-.002	.006	-.004	.002	-.001	.001	.002	-.005	-.001
13	-.003	-.005	.004	-.005	-.001	-.004	.003	.002	-.001	-.002
14	-.001	-.005	.001	-.001	-.003	-.005	-.001	.003	.001	-.000
15	-.005	-.005	.002	-.005	.001	-.006	-.002	.004	-.003	.001
16	-.007	-.002	.002	-.006	-.000	-.006	-.004	.005	-.002	.003
17	-.009	-.001	.002	-.003	.001	-.005	-.003	.006	-.003	.004
18	-.009	-.000	.001	-.002	-.002	-.006	-.002	.004	.001	.005
19	-.005	-.000	-.001	-.002	-.002	-.006	-.000	.001	.002	.003
20	-.004	-.002	.001	-.001	-.004	-.007	-.002	.004	-.000	.002
21	-.003	-.000	.001	.001	-.001	-.004	-.002	.005	.003	.001
22	-.001	.002	.003	.006	.003	-.001	.001	.007	.003	.002
23	.004	.007	.007	.019	.009	.007	.005	.013	.008	.006
24	.038	.016	.015	.029	.013	.018	.013	.026	.017	.011
25	.013	.044	.023	.037	.023	.028	.022	.043	.037	.019
26	.027	.070	.038	.040	.030	.025	.025	.040	.034	.029
27	.047	.049	.049	.040	.038	.034	.029	.036	.046	.037
28	.064	.040	.055	.041	.041	.040	.039	.036	.044	.043
29	.058	.051	.053	.042	.044	.050	.046	.034	.053	.046
30	.069	.050	.053	.047	.050	.053	.050	.041	.059	.043
31	.063	.045	.052	.046	.052	.054	.048	.045	.058	.043
32	.066	.043	.051	.049	.050	.060	.047	.048	.054	.043
33	.059	.046	.054	.046	.051	.057	.049	.049	.056	.045
34	.056	.048	.049	.044	.052	.059	.046	.051	.055	.045
35	.059	.049	.049	.047	.050	.054	.047	.051	.053	.044
36	.058	.052	.049	.049	.051	.053	.051	.054	.054	.041
37	.057	.048	.047	.048	.049	.054	.053	.053	.052	.040
38	.056	.046	.047	.050	.053	.051	.050	.049	.054	.044
39	.055	.048	.047	.052	.052	.047	.048	.048	.054	.044
40	.053	.050	.048	.057	.057	.051	.050	.050	.056	.047
41	.054	.050	.046	.057	.053	.052	.048	.049	.056	.046
42	.053	.049	.049	.051	.055	.053	.048	.048	.053	.046
43	.054	.049	.048	.049	.052	.058	.047	.047	.051	.049
44	.055	.052	.048	.046	.051	.055	.045	.050	.048	.045
45	.054	.052	.049	.051	.050	.058	.042	.049	.048	.046
46	.052	.052	.050	.053	.051	.053	.046	.050	.043	.044
47	.053	.053	.050	.054	.049	.056	.045	.052	.050	.044
48	.052	.053	.048	.055	.049	.056	.045	.052	.049	.048
49	.052	.054	.051	.051	.049	.055	.045	.053	.046	.057
50	.053	.054	.050	.050	.046	.055	.045	.056	.046	.048

b_t (estimate of β_t) produced by MSM to a growth change of size 1σ at $t = 21$

i.e. $\beta_t = \begin{cases} 0 & t < 21 \\ 0.050 & t \geqslant 21 \end{cases}$

APPENDIX M cont.

TABLE M.1c

t	REALISATION 1	2	3	4	5	6	7	8	9	10
1	.004	-.003	-.001	.006	.000	.002	-.002	-.004	-.003	-.006
2	.000	.003	-.008	-.015	.009	.012	.002	-.001	-.005	-.001
3	.001	.012	.005	.012	.010	-.006	-.001	.020	.009	-.007
4	-.006	.008	.009	.008	.011	-.015	-.000	.017	.006	.002
5	.010	-.007	.014	-.001	.004	-.020	.004	.008	.000	.003
6	.007	-.003	.013	-.011	.003	-.017	.012	.001	.004	.008
7	.007	-.009	.014	-.008	-.005	-.014	.007	.004	-.006	.010
8	.005	-.008	.007	-.016	-.007	-.007	.002	.009	-.008	.004
9	.006	-.004	.008	-.009	-.001	-.000	.007	.010	-.006	.003
10	.000	-.001	.008	-.006	.001	-.003	.005	.004	-.006	.006
11	-.002	-.004	.005	-.007	.001	.005	.006	.005	-.001	.002
12	-.001	-.002	.006	-.004	.002	-.001	.001	.002	-.005	-.001
13	-.003	-.005	.004	-.005	-.001	-.004	.003	.002	-.001	-.002
14	-.001	-.005	.001	-.001	-.003	-.005	-.001	.003	.001	-.000
15	-.005	-.005	.002	-.005	.001	-.006	-.002	.004	-.003	.001
16	-.007	-.002	.002	-.006	-.000	-.006	-.004	.005	-.002	.003
17	-.009	-.001	.002	-.003	.001	-.005	-.003	.006	-.003	.004
18	-.009	-.000	.001	-.002	-.002	-.006	-.002	.004	.001	.005
19	-.005	-.000	-.001	-.002	-.002	-.006	-.000	.001	.002	.003
20	-.004	-.002	.001	-.001	-.004	-.007	-.002	.004	-.000	.002
21	-.002	.000	.001	.001	-.000	-.004	-.002	.006	.004	.002
22	.002	.005	.005	.007	.007	.004	-.000	.010	.006	.004
23	.007	.021	.016	.041	.030	.034	.014	.019	.009	.009
24	.074	.051	.051	.056	.040	.052	.029	.059	.031	.040
25	.032	.073	.056	.067	.055	.060	.043	.092	.062	.056
26	.049	.093	.076	.067	.061	.051	.043	.068	.050	.068
27	.073	.076	.081	.065	.068	.063	.049	.057	.071	.073
28	.092	.068	.084	.066	.069	.068	.065	.056	.067	.077
29	.082	.077	.080	.067	.070	.078	.072	.053	.080	.076
30	.094	.075	.079	.073	.076	.079	.074	.064	.087	.070
31	.087	.071	.077	.072	.077	.079	.072	.069	.085	.068
32	.090	.070	.076	.075	.075	.083	.071	.073	.080	.068
33	.084	.072	.079	.071	.075	.081	.072	.074	.081	.069
34	.080	.073	.073	.068	.076	.082	.070	.076	.080	.069
35	.083	.074	.074	.071	.074	.078	.071	.075	.078	.068
36	.082	.076	.073	.074	.075	.077	.075	.079	.078	.065
37	.082	.073	.072	.073	.074	.078	.077	.077	.077	.064
38	.081	.071	.072	.075	.078	.075	.075	.074	.078	.068
39	.080	.073	.071	.077	.076	.072	.072	.073	.079	.068
40	.078	.075	.073	.082	.081	.076	.074	.074	.081	.071
41	.078	.075	.071	.082	.077	.076	.072	.073	.080	.070
42	.078	.074	.074	.076	.079	.078	.072	.073	.078	.071
43	.079	.074	.073	.074	.076	.083	.071	.072	.076	.073
44	.079	.077	.072	.071	.076	.080	.069	.075	.073	.070
45	.079	.077	.074	.076	.075	.083	.067	.074	.073	.070
46	.077	.077	.075	.077	.075	.078	.071	.075	.068	.069
47	.078	.078	.074	.078	.073	.081	.069	.077	.075	.069
48	.076	.078	.073	.080	.074	.081	.070	.077	.074	.073
49	.077	.079	.075	.075	.074	.080	.070	.078	.071	.082
50	.078	.079	.075	.075	.071	.080	.070	.081	.071	.073

b_t (estimate of β_t) produced by MSM to a growth change of size 1.5σ at $t = 21$

$$\text{i.e. } \beta_t = \begin{cases} 0 & t < 21 \\ 0.075 & t \geq 21 \end{cases}$$

TABLE M.2a

	REALISATION									
t	1	2	3	4	5	6	7	8	9	10
1	.004	-.003	-.001	.006	.000	.002	-.002	-.004	-.003	-.006
2	.000	.003	-.008	-.015	.009	.012	.002	-.001	-.005	-.001
3	.001	.012	.005	.012	.010	-.006	-.001	.020	.009	-.007
4	-.006	.008	.009	.008	.011	-.015	-.000	.017	.006	.002
5	.010	-.007	.014	-.000	.004	-.020	.004	.008	.000	.003
6	.007	-.003	.013	-.011	.003	-.017	.012	.001	.004	.008
7	.007	-.009	.014	-.007	-.005	-.014	.007	.004	-.006	.010
8	.005	-.008	.007	-.016	-.007	-.007	.002	.009	-.008	.004
9	.006	-.004	.008	-.009	-.001	-.000	.007	.010	-.006	.003
10	.000	-.001	.008	-.006	.001	-.003	.005	.004	-.006	.006
11	-.002	-.004	.005	-.007	.001	.004	.006	.005	-.001	.002
12	-.000	-.002	.006	-.004	.001	-.001	.001	.002	-.005	-.001
13	-.003	-.005	.004	-.005	-.001	-.004	.003	.002	-.001	-.001
14	-.001	-.005	.002	-.001	-.003	-.005	-.000	.003	.000	.000
15	-.004	-.005	.002	-.005	.001	-.005	-.002	.004	-.003	.001
16	-.007	-.002	.002	-.005	-.000	-.006	-.003	.005	-.002	.003
17	-.008	-.001	.003	-.004	.001	-.005	-.003	.006	-.002	.004
18	-.008	-.001	.001	-.003	-.002	-.006	-.002	.004	.001	.004
19	-.005	-.001	-.001	-.002	-.002	-.006	-.000	.002	.001	.002
20	-.004	-.002	.001	-.002	-.003	-.006	-.002	.004	-.001	.002
21	-.004	-.001	.000	-.001	-.001	-.005	-.003	.004	.001	.001
22	-.003	-.000	.001	.002	-.000	-.004	.000	.004	.000	.000
23	.001	.001	.002	.004	.001	-.002	-.000	.007	.003	.002
24	.040	.009	.004	.007	.001	-.000	.001	.008	.005	.003
25	.003	.016	.004	.010	.004	.002	.003	.015	.010	.004
26	.008	.035	.008	.012	.005	.001	.003	.011	.008	.005
27	.016	.024	.014	.013	.009	.005	.004	.010	.016	.007
28	.026	.019	.023	.014	.011	.009	.009	.010	.017	.010
29	.024	.026	.023	.016	.014	.034	.017	.009	.024	.013
30	.032	.025	.025	.019	.021	.036	.024	.015	.030	.013
31	.030	.022	.025	.019	.026	.035	.022	.018	.030	.013
32	.033	.021	.025	.022	.024	.043	.021	.021	.027	.015
33	.030	.022	.029	.020	.025	.036	.024	.022	.029	.017
34	.029	.023	.023	.018	.027	.037	.020	.025	.029	.017
35	.031	.024	.024	.021	.024	.029	.023	.025	.027	.017
36	.031	.026	.023	.023	.026	.027	.028	.029	.028	.015
37	.031	.023	.022	.022	.025	.028	.029	.027	.027	.015
38	.030	.022	.022	.024	.029	.025	.026	.024	.028	.018
39	.030	.024	.022	.026	.027	.020	.023	.023	.029	.018
40	.028	.025	.023	.033	.032	.025	.026	.025	.030	.020
41	.028	.025	.021	.031	.028	.025	.024	.024	.030	.020
42	.028	.024	.024	.025	.030	.026	.023	.024	.028	.020
43	.029	.024	.023	.024	.027	.031	.022	.023	.027	.023
44	.029	.027	.023	.022	.026	.029	.021	.025	.024	.020
45	.029	.026	.024	.026	.025	.031	.019	.024	.024	.020
46	.028	.026	.025	.028	.026	.028	.022	.025	.018	.019
47	.028	.027	.024	.029	.024	.030	.021	.026	.026	.019
48	.027	.027	.023	.030	.024	.030	.020	.026	.024	.023
49	.028	.027	.025	.026	.024	.029	.021	.027	.021	.039
50	.028	.027	.024	.026	.022	.029	.020	.030	.021	.023

b_t (estimate of β_t) produced by 2PS to a growth change of size ½ σ at $t = 21$

$$\text{i.e. } \beta_t = \begin{cases} 0 & t < 21 \\ 0.025 & t \geq 21 \end{cases}$$

APPENDIX M cont.

TABLE M.2b

t	REALISATION 1	2	3	4	5	6	7	8	9	10
1	.004	-.003	-.001	.006	.000	.002	-.002	-.004	-.003	-.006
2	.000	.003	-.008	-.015	.009	.012	.002	-.001	-.005	-.001
3	.001	.012	.005	.012	.010	-.006	-.001	.020	.009	-.007
4	-.006	.008	.009	.008	.011	-.015	-.000	.017	.006	.002
5	.010	-.007	.014	-.000	.004	-.020	.004	.008	.000	.003
6	.007	-.003	.013	-.011	.003	-.017	.012	.001	.004	.008
7	.007	-.009	.014	-.007	-.005	-.014	.007	.004	-.006	.010
8	.005	-.008	.007	-.016	-.007	-.007	.002	.009	-.008	.004
9	.006	-.004	.008	-.009	-.001	-.000	.007	.010	-.006	.003
10	.000	-.001	.008	-.006	.001	-.003	.005	.004	-.006	.006
11	-.002	-.004	.005	-.007	.001	.004	.006	.005	-.001	.002
12	-.000	-.002	.006	-.004	.001	-.001	.001	.002	-.005	-.001
13	-.003	-.005	.004	-.005	-.001	-.004	.003	.002	-.001	-.001
14	-.001	-.005	.002	-.001	-.003	-.005	-.000	.003	.000	.000
15	-.004	-.005	.002	-.005	.001	-.005	-.002	.004	-.003	.001
16	-.007	-.002	.002	-.005	-.000	-.006	-.003	.005	-.002	.003
17	-.008	-.001	.003	-.004	.001	-.005	-.003	.006	-.002	.004
18	-.008	-.001	.001	-.003	-.002	-.006	-.002	.004	.001	.004
19	-.005	-.001	-.001	-.002	-.002	-.006	-.000	.002	.001	.002
20	-.004	-.002	.001	-.002	-.003	-.006	-.002	.004	-.001	.002
21	-.004	-.001	.001	-.000	-.001	-.004	-.002	.004	.002	.001
22	-.001	.001	.003	.006	.002	-.001	.001	.006	.002	.002
23	.005	.007	.006	.023	.009	.009	.003	.013	.006	.005
24	.076	.044	.021	.034	.016	.028	.011	.030	.024	.010
25	.024	.056	.030	.042	.030	.036	.027	.060	.050	.022
26	.021	.072	.050	.042	.035	.028	.027	.043	.034	.036
27	.038	.055	.055	.041	.042	.038	.031	.036	.049	.043
28	.056	.045	.058	.042	.043	.043	.044	.035	.045	.048
29	.052	.052	.054	.043	.044	.052	.048	.033	.055	.049
30	.063	.051	.053	.048	.050	.053	.050	.040	.061	.045
31	.059	.047	.052	.047	.051	.053	.048	.044	.059	.043
32	.062	.046	.051	.049	.049	.058	.047	.047	.055	.043
33	.057	.047	.054	.046	.050	.056	.048	.048	.056	.045
34	.055	.048	.049	.044	.051	.057	.045	.050	.055	.045
35	.057	.049	.049	.046	.049	.053	.047	.050	.053	.043
36	.057	.051	.049	.049	.050	.052	.050	.053	.054	.041
37	.056	.048	.047	.048	.049	.053	.051	.052	.052	.040
38	.056	.047	.047	.049	.052	.050	.050	.049	.054	.044
39	.055	.048	.047	.051	.051	.047	.047	.048	.054	.043
40	.053	.049	.048	.057	.055	.051	.049	.049	.055	.045
41	.054	.050	.046	.055	.052	.051	.048	.048	.055	.045
42	.053	.049	.049	.050	.053	.052	.048	.048	.053	.045
43	.054	.049	.048	.049	.051	.056	.047	.048	.052	.047
44	.054	.051	.048	.047	.050	.054	.045	.050	.049	.045
45	.054	.051	.048	.051	.049	.056	.044	.049	.049	.045
46	.053	.051	.049	.052	.050	.052	.047	.050	.043	.044
47	.053	.052	.049	.053	.048	.055	.045	.051	.051	.044
48	.052	.052	.048	.054	.048	.055	.045	.051	.049	.047
49	.053	.052	.050	.051	.049	.054	.046	.052	.046	.063
50	.053	.052	.049	.050	.046	.054	.045	.055	.046	.048

b_t (estimate of β_t) produced by 2PS to a growth change of size 1σ at $t = 21$

i.e. $$\beta_t = \begin{cases} 0 & t < 21 \\ 0.050 & t \geq 21 \end{cases}$$

APPENDIX M cont.

TABLE M.2c

t	REALISATION 1	2	3	4	5	6	7	8	9	10
1	.004	-.003	-.001	.006	.000	.002	-.002	-.004	-.003	-.006
2	.000	.003	-.008	-.015	.009	.012	.002	-.001	-.005	-.001
3	.001	.012	.005	.012	.010	-.006	-.001	.020	.009	-.007
4	-.006	.008	.009	.008	.011	-.015	-.000	.017	.006	.002
5	.010	-.007	.014	-.000	.004	-.020	.004	.008	.000	.003
6	.007	-.003	.013	-.011	.003	-.017	.012	.001	.004	.008
7	.007	-.009	.014	-.007	-.005	-.014	.007	.004	-.006	.010
8	.005	-.008	.007	-.016	-.007	-.007	.002	.009	-.008	.004
9	.006	-.004	.008	-.009	-.001	-.000	.007	.010	-.006	.003
10	.000	-.001	.008	-.006	.001	-.003	.005	.004	-.006	.006
11	-.002	-.004	.005	-.007	.001	.004	.006	.005	-.001	.002
12	-.000	-.002	.006	-.004	.001	-.001	.001	.002	-.005	-.001
13	-.003	-.005	.004	-.005	-.001	-.004	.003	.002	-.001	-.001
14	-.001	-.005	.002	-.001	-.003	-.005	-.000	.003	.000	.000
15	-.004	-.005	.002	-.005	.001	-.005	-.002	.004	-.003	.001
16	-.007	-.002	.002	-.005	-.000	-.006	-.003	.005	-.002	.003
17	-.008	-.001	.003	-.004	.001	-.005	-.003	.006	-.002	.004
18	-.008	-.001	.001	-.003	-.002	-.006	-.002	.004	.001	.004
19	-.005	-.001	-.001	-.002	-.002	-.006	-.000	.002	.001	.002
20	-.004	-.002	.001	-.002	-.003	-.006	-.002	.004	-.001	.002
21	-.003	-.000	.001	.000	-.000	-.004	-.001	.005	.002	.002
22	.001	.005	.005	.010	.008	.006	-.000	.008	.004	.003
23	.013	.044	.024	.053	.046	.049	.011	.031	.014	.010
24	.106	.093	.069	.059	.046	.063	.029	.069	.063	.049
25	.040	.084	.062	.068	.064	.065	.057	.095	.080	.059
26	.049	.092	.081	.067	.066	.052	.052	.075	.065	.071
27	.071	.081	.082	.065	.072	.063	.056	.064	.073	.074
28	.086	.073	.085	.066	.071	.068	.072	.062	.069	.077
29	.080	.078	.080	.067	.071	.078	.075	.059	.078	.075
30	.089	.077	.079	.072	.077	.078	.076	.066	.083	.070
31	.085	.073	.077	.071	.077	.078	.073	.069	.081	.068
32	.087	.072	.076	.074	.075	.082	.072	.072	.078	.068
33	.083	.073	.079	.071	.075	.080	.073	.072	.079	.069
34	.080	.074	.074	.068	.076	.081	.070	.074	.079	.069
35	.082	.074	.074	.071	.074	.077	.072	.074	.077	.068
36	.082	.076	.074	.073	.075	.077	.075	.077	.078	.065
37	.081	.073	.072	.072	.074	.077	.076	.076	.077	.064
38	.081	.072	.072	.074	.077	.075	.074	.074	.078	.068
39	.080	.073	.072	.076	.076	.072	.072	.073	.079	.068
40	.078	.074	.073	.082	.080	.075	.074	.074	.080	.070
41	.079	.075	.071	.080	.076	.076	.073	.073	.080	.069
42	.078	.074	.074	.075	.078	.077	.073	.073	.078	.070
43	.079	.074	.073	.073	.075	.081	.072	.072	.077	.072
44	.079	.076	.073	.072	.075	.079	.070	.075	.074	.070
45	.079	.076	.073	.076	.074	.081	.069	.074	.074	.070
46	.078	.076	.074	.077	.075	.077	.072	.074	.068	.069
47	.078	.076	.074	.078	.073	.080	.070	.076	.076	.069
48	.077	.077	.073	.079	.073	.080	.070	.076	.074	.072
49	.078	.077	.075	.076	.074	.079	.071	.077	.071	.089
50	.078	.077	.074	.075	.071	.079	.070	.079	.071	.073

b_t (estimate of β_t) produced by 2PS to a growth change of size 1.5 σ at $t=21$

i.e. $\beta_t = \begin{cases} 0 & t < 21 \\ 0.075 & t \geqslant 21 \end{cases}$

APPENDIX N

TABLE N.1

t	$\rho^{(11)}$	$\rho^{(21)}$	$\rho^{(31)}$	$\rho^{(41)}$	$\rho^{(12)}$	$\rho^{(22)}$	$\rho^{(32)}$	$\rho^{(42)}$	$\rho^{(13)}$	$\rho^{(23)}$	$\rho^{(33)}$	$\rho^{(43)}$	$\rho^{(14)}$	$\rho^{(24)}$	$\rho^{(34)}$	$\rho^{(44)}$	33S b_t	MSM b_t
1	876	91	2	3	21	2	0	0	3	0	1	0	1	0	0	0	.000	.000
2	959	15	2	1	18	0	0	0	3	0	1	0	1	0	0	0	.009	.009
3	965	13	2	1	15	0	0	0	3	0	1	0	0	0	0	0	.010	.010
4	967	12	2	0	15	0	0	0	3	0	0	0	0	0	0	0	.011	.011
5	962	13	2	0	18	0	0	0	3	0	1	0	1	0	0	0	.004	.004
6	967	13	2	1	14	0	0	0	3	0	0	0	0	0	0	0	.003	.003
7	952	13	3	0	27	0	0	0	3	0	1	0	1	0	0	0	-.005	-.005
8	962	16	2	1	14	0	0	0	3	0	1	0	0	0	0	0	-.007	-.007
9	933	20	2	0	38	1	0	0	4	0	1	0	1	0	0	0	-.001	-.001
10	955	23	2	1	14	1	0	0	3	0	1	0	0	0	0	0	.001	.001
11	970	13	1	0	12	0	0	0	2	0	0	0	0	0	0	0	.001	.001
12	971	11	1	0	12	0	0	0	2	0	0	0	0	0	0	0	.001	.002
13	963	12	2	0	18	0	0	0	3	0	0	0	1	0	0	0	-.001	-.001
14	957	13	3	1	21	0	0	0	3	0	1	0	1	0	0	0	-.003	-.003
15	876	39	1	0	74	2	0	0	5	0	1	0	2	0	0	0	.001	.001
16	902	79	1	1	12	1	0	0	2	0	0	0	0	0	0	0	.000	-.000
17	966	13	2	0	14	0	0	0	3	0	0	0	0	0	0	0	.001	.001
18	919	20	2	0	51	1	0	0	4	0	1	0	2	0	0	0	-.002	-.002
19	941	41	2	1	11	1	0	0	2	0	0	0	0	0	0	0	-.002	-.002
20	958	11	3	0	22	0	0	0	3	0	1	0	1	0	0	0	-.004	-.004
21	0	0	0	0	943	23	3	1	0	0	0	0	30	1	0	0	-.004	-.004
22	0	984	0	0	0	13	0	0	0	3	0	0	0	0	0	0	-.003	-.003
23	975	8	1	0	12	0	0	0	3	0	0	0	0	0	0	0	-.003	-.002
24	963	13	2	0	17	0	0	0	3	0	1	0	1	0	0	0	-.004	-.004
25	964	19	0	0	13	0	0	0	3	0	0	0	0	0	0	0	-.003	-.003
26	970	13	1	0	12	0	0	0	2	0	0	0	0	0	0	0	-.004	-.003
27	970	12	1	0	13	0	0	0	3	0	0	0	0	0	0	0	-.002	-.003
28	970	13	1	0	12	0	0	0	2	0	0	0	0	0	0	0	-.003	-.003
29	972	11	1	0	12	0	0	0	2	0	0	0	0	0	0	0	-.003	-.003
30	947	12	2	0	32	0	0	0	4	0	1	0	1	0	0	0	-.000	-.001
31	956	24	2	1	13	0	0	0	3	0	1	0	0	0	0	0	.001	-.000
32	966	13	1	0	15	0	0	0	3	0	0	0	0	0	0	0	-.001	-.001
33	968	15	1	0	12	0	0	0	2	0	0	0	0	0	0	0	-.001	-.001
34	971	11	1	0	13	0	0	0	3	0	0	0	0	0	0	0	.000	-.000
35	961	14	2	0	19	0	0	0	3	0	0	0	1	0	0	0	-.001	-.002
36	959	21	1	0	14	0	0	0	3	0	0	0	0	0	0	0	-.000	-.001
37	962	16	2	0	16	0	0	0	3	0	0	0	1	0	0	0	-.002	-.002
38	710	23	4	0	239	4	1	0	8	0	2	0	8	0	0	0	.002	.001
39	721	261	1	1	10	3	0	0	2	1	0	0	0	0	0	0	.001	.000
40	673	23	10	0	269	4	1	0	8	0	3	0	9	0	0	0	.005	.005
41	0	1	0	0	675	278	9	7	0	0	0	0	22	9	0	0	.005	.004
42	11	224	1	378	0	357	0	12	0	3	0	1	0	11	0	0	.001	.000
43	896	4	13	14	42	21	0	1	3	0	3	0	1	1	0	0	.003	.002
44	960	19	1	2	13	1	0	0	3	0	0	0	0	0	0	0	.003	.003
45	970	11	2	0	14	0	0	0	3	0	0	0	0	0	0	0	.002	.001
46	968	13	2	0	14	0	0	0	3	0	0	0	0	0	0	0	.004	.004
47	957	16	2	0	21	0	0	0	3	0	0	0	1	0	0	0	-.002	-.003
48	963	18	2	1	13	0	0	0	3	0	0	0	0	0	0	0	-.001	-.002
49	969	12	2	0	13	0	0	0	3	0	0	0	0	0	0	0	-.000	-.000
50	935	17	2	0	39	1	0	0	4	0	1	0	1	0	0	0	-.007	-.007

$\rho^{(ij)} \times 10^3$ and b_t (estimate of β_t) produced by 33S when applied to C8 Data with $\pi^{(33)} = \cdot 2$ and using the standard SSP. The last column is the MSM b_t for the same data

TABLE N.1 cont.

51	951	30	2	1	12	1	0	0	2	0	0	0	0	0	0	0	-.007	-.007
52	943	14	2	0	35	0	0	0	4	0	1	0	1	0	0	0	-.002	-.002
53	909	57	1	0	26	1	0	0	3	0	0	0	1	0	0	0	-.006	-.006
54	961	19	2	1	13	0	0	0	3	0	1	0	0	0	0	0	-.007	-.007
55	969	12	2	0	13	0	0	0	3	0	0	0	0	0	0	0	-.008	-.008
56	943	17	2	0	32	0	0	0	4	0	1	0	1	0	0	0	-.005	-.005
57	956	22	2	1	14	0	0	0	3	0	1	0	0	0	0	0	-.004	-.004
58	943	12	5	1	34	1	0	0	4	0	1	0	1	0	0	0	-.001	-.001
59	933	44	1	0	17	1	0	0	3	0	0	0	1	0	0	0	-.003	-.003
60	964	14	2	1	15	0	0	0	3	0	1	0	0	0	0	0	-.005	-.004
61	948	19	1	0	26	0	0	0	3	0	0	0	1	0	0	0	-.002	-.002
62	950	17	4	1	22	1	0	0	3	0	1	0	1	0	0	0	.000	.000
63	954	16	4	1	21	0	0	0	3	0	1	0	1	0	0	0	.002	.002
64	145	2	52	3	732	16	3	1	5	0	16	0	23	1	0	0	.008	.006
65	623	35	68	158	19	68	2	2	2	1	18	0	1	2	0	0	.036	.033
66	938	12	18	3	18	2	0	0	3	0	5	0	1	0	0	0	.046	.046
67	941	13	8	1	31	1	0	0	3	0	2	0	1	0	0	0	.061	.063
68	903	56	2	1	33	1	0	0	3	0	1	0	1	0	0	0	.047	.048
69	929	43	2	1	21	1	0	0	3	0	1	0	1	0	0	0	.055	.057
70	833	59	3	0	93	2	0	0	5	0	1	0	3	0	0	0	.042	.042
71	903	76	2	1	13	1	0	0	2	0	1	0	0	0	0	0	.044	.044
72	963	12	2	0	18	0	0	0	3	0	1	0	1	0	0	0	.048	.048
73	965	13	2	1	15	0	0	0	3	0	1	0	0	0	0	0	.050	.051
74	967	14	1	0	13	0	0	0	3	0	0	0	0	0	0	0	.049	.049
75	969	12	2	0	14	0	0	0	3	0	0	0	0	0	0	0	.048	.048
76	970	13	1	0	12	0	0	0	2	0	0	0	0	0	0	0	.048	.048
77	963	12	2	0	19	0	0	0	3	0	0	0	1	0	0	0	.050	.050
78	951	24	1	0	19	0	0	0	3	0	0	0	1	0	0	0	.048	.048
79	963	19	1	0	12	0	0	0	2	0	0	0	0	0	0	0	.048	.049
80	966	12	2	0	15	0	0	0	3	0	0	0	0	0	0	0	.047	.047
81	881	23	3	0	83	1	0	0	5	0	1	0	3	0	0	0	.050	.051
82	923	53	3	2	13	1	0	0	3	0	1	0	0	0	0	0	.051	.052
83	969	13	2	0	13	0	0	0	3	0	0	0	0	0	0	0	.052	.052
84	971	13	1	0	12	0	0	0	2	0	0	0	0	0	0	0	.052	.052
85	956	11	2	0	25	0	0	0	3	0	1	0	1	0	0	0	.050	.050
86	951	29	1	0	15	0	0	0	3	0	0	0	0	0	0	0	.051	.051
87	957	18	2	0	19	0	0	0	3	0	0	0	1	0	0	0	.049	.049
88	959	21	1	0	14	0	0	0	3	0	0	0	0	0	0	0	.050	.051
89	937	18	2	0	35	1	0	0	4	0	1	0	1	0	0	0	.048	.048
90	947	35	1	1	12	0	0	0	2	0	0	0	0	0	0	0	.049	.049
91	971	12	1	0	12	0	0	0	2	0	0	0	0	0	0	0	.048	.048
92	966	12	2	0	16	0	0	0	3	0	0	0	1	0	0	0	.049	.049
93	957	19	1	0	18	0	0	0	3	0	0	0	1	0	0	0	.048	.048
94	922	26	2	0	42	1	0	0	4	0	1	0	1	0	0	0	.050	.050
95	927	52	1	0	15	1	0	0	3	0	0	0	0	0	0	0	.049	.049
96	965	13	2	1	15	0	0	0	3	0	1	0	0	0	0	0	.048	.048
97	969	15	1	0	11	0	0	0	2	0	0	0	0	0	0	0	.048	.048
98	971	11	1	0	13	0	0	0	3	0	0	0	0	0	0	0	.049	.049
99	968	14	1	0	13	0	0	0	3	0	0	0	0	0	0	0	.048	.048
100	970	12	2	0	13	0	0	0	3	0	0	0	0	0	0	0	.047	.047

APPENDIX N cont.

TABLE N.2

t	$p^{(11)}$	$p^{(21)}$	$p^{(31)}$	$p^{(41)}$	$p^{(12)}$	$p^{(22)}$	$p^{(32)}$	$p^{(42)}$	$p^{(13)}$	$p^{(23)}$	$p^{(33)}$	$p^{(43)}$	$p^{(14)}$	$p^{(24)}$	$p^{(34)}$	$p^{(44)}$	33S b_t	MSM b_t
1	876	91	1	3	21	2	0	0	3	0	2	0	1	0	0	0	.000	.000
2	958	15	1	1	18	0	0	0	3	0	2	0	1	0	0	0	.009	.009
3	964	13	1	1	15	0	0	0	3	0	2	0	0	0	0	0	.010	.010
4	966	12	1	0	15	0	0	0	3	0	2	0	0	0	0	0	.011	.011
5	961	13	1	0	18	0	0	0	3	0	2	0	1	0	0	0	.004	.004
6	965	13	1	1	14	0	0	0	3	0	2	0	0	0	0	0	.003	.003
7	950	13	2	0	27	0	0	0	3	0	3	0	1	0	0	0	-.005	-.005
8	960	17	2	1	14	0	0	0	3	0	3	0	0	0	0	0	-.007	-.007
9	931	20	1	0	38	1	0	0	4	0	3	0	1	0	0	0	-.001	-.001
10	954	23	2	1	14	1	0	0	3	0	3	0	0	0	0	0	.001	.001
11	969	13	1	0	12	0	0	0	2	0	2	0	0	0	0	0	.001	.001
12	970	11	1	0	12	0	0	0	2	0	2	0	0	0	0	0	.001	.002
13	962	12	1	0	18	0	0	0	3	0	2	0	1	0	0	0	-.001	-.001
14	955	13	2	1	21	0	0	0	3	0	3	0	1	0	0	0	-.003	-.003
15	874	38	1	0	73	2	0	0	5	0	3	0	2	0	0	0	.001	.001
16	902	79	1	1	12	1	0	0	2	0	2	0	0	0	0	0	-.000	-.000
17	966	13	1	0	14	0	0	0	3	0	2	0	0	0	0	0	.001	.001
18	918	20	1	0	51	1	0	0	4	0	3	0	2	0	0	0	-.002	-.002
19	940	41	1	1	11	1	0	0	2	0	2	0	0	0	0	0	-.002	-.002
20	956	12	2	0	22	0	0	0	3	0	3	0	1	0	0	0	-.004	-.004
21	0	0	0	0	940	23	5	1	0	0	0	0	30	1	0	0	-.004	-.004
22	0	984	0	0	0	13	0	0	0	3	0	0	0	0	0	0	-.003	-.003
23	975	8	0	0	12	0	0	0	3	0	1	0	0	0	0	0	-.003	-.002
24	962	13	1	0	18	0	0	0	3	0	2	0	1	0	0	0	-.004	-.004
25	964	19	0	0	13	0	0	0	3	0	1	0	0	0	0	0	-.003	-.003
26	970	13	1	0	12	0	0	0	2	0	1	0	0	0	0	0	-.003	-.003
27	969	12	1	0	13	0	0	0	3	0	1	0	0	0	0	0	-.002	-.003
28	970	13	1	0	12	0	0	0	2	0	1	0	0	0	0	0	-.003	-.003
29	972	11	1	0	12	0	0	0	2	0	1	0	0	0	0	0	-.003	-.003
30	946	12	1	0	32	0	0	0	4	0	3	0	1	0	0	0	-.000	-.001
31	954	25	1	1	13	0	0	0	3	0	2	0	0	0	0	0	.000	-.000
32	966	13	1	0	15	0	0	0	3	0	2	0	0	0	0	0	-.001	-.001
33	969	14	1	0	11	0	0	0	2	0	1	0	0	0	0	0	-.001	-.001
34	970	11	1	0	13	0	0	0	3	0	1	0	0	0	0	0	-.000	-.000
35	961	14	1	0	18	0	0	0	3	0	2	0	1	0	0	0	-.002	-.002
36	960	20	1	0	14	0	0	0	3	0	1	0	0	0	0	0	-.001	-.001
37	962	16	1	0	15	0	0	0	3	0	2	0	0	0	0	0	-.002	-.002
38	693	21	3	0	254	4	1	0	8	0	8	0	8	0	0	0	.002	.001
39	709	272	1	1	9	3	0	0	2	1	2	0	0	0	0	0	.000	.000
40	622	25	7	1	305	4	1	0	8	0	17	0	10	0	0	0	.005	.005
41	0	1	0	0	627	314	19	8	0	0	0	0	20	10	1	0	.004	.004
42	9	329	0	325	0	310	0	11	0	3	1	1	0	10	0	0	-.002	.000
43	881	6	8	14	51	21	0	1	3	0	13	0	2	1	0	0	.000	.002
44	954	22	2	2	13	1	0	0	3	0	3	0	0	0	0	0	.001	.003
45	968	10	1	0	14	0	0	0	3	0	2	0	0	0	0	0	-.001	.001
46	966	13	1	0	14	0	0	0	3	0	2	0	0	0	0	0	.002	.004
47	957	16	1	0	20	0	0	0	3	0	2	0	1	0	0	0	-.004	-.003
48	962	18	1	0	13	0	0	0	3	0	2	0	0	0	0	0	-.003	-.002
49	968	12	1	0	13	0	0	0	3	0	2	0	0	0	0	0	-.001	-.000
50	936	17	1	0	36	1	0	0	4	0	3	0	1	0	0	0	-.008	-.007

$p^{(ij)} \times 10^3$ and b_t (estimate of β_t) produced by 33S when applied to CB Data with $\pi^{(33)} = \cdot 6$ and using the standard SSP

The last column is the MSM b_t for the same data

APPENDIX N cont.

TABLE N.2 cont.

51	952	29	1	1	12	0	0	0	2	0	2	0	0	0	0	0	-.007	-.007
52	939	13	2	0	37	0	0	0	4	0	3	0	1	0	0	0	-.002	-.002
53	907	60	1	0	25	1	0	0	3	0	2	0	1	0	0	0	-.006	-.006
54	961	18	1	1	13	0	0	0	3	0	2	0	0	0	0	0	-.007	-.007
55	968	12	1	0	13	0	0	0	3	0	2	0	0	0	0	0	-.008	-.008
56	941	16	1	0	33	0	0	0	4	0	3	0	1	0	0	0	-.005	-.005
57	954	23	2	1	14	1	0	0	3	0	3	0	0	0	0	0	-.004	-.004
58	937	12	3	1	35	1	0	0	4	0	6	0	1	0	0	0	-.002	-.001
59	931	45	1	0	16	1	0	0	3	0	2	0	1	0	0	0	-.004	-.003
60	964	14	1	1	15	0	0	0	3	0	2	0	0	0	0	0	-.005	-.004
61	946	19	1	0	27	0	0	0	3	0	2	0	1	0	0	0	-.002	-.002
62	947	17	2	1	23	1	0	0	3	0	4	0	1	0	0	0	-.000	.000
63	948	16	3	1	22	1	0	0	3	0	5	0	1	0	0	0	.002	.002
64	99	2	44	3	705	16	6	1	4	0	96	0	23	1	0	0	.014	.006
65	408	24	122	136	12	58	9	2	1	1	224	0	0	2	0	0	.042	.033
66	731	21	81	3	14	1	5	0	2	0	142	0	0	0	0	0	.049	.046
67	799	14	55	1	23	1	4	0	3	0	101	0	1	0	0	0	.066	.063
68	832	50	21	1	36	1	4	0	3	0	50	0	1	0	0	0	.046	.048
69	897	46	10	1	20	1	1	0	3	0	20	0	1	0	0	0	.056	.057
70	801	64	9	0	91	2	2	0	5	0	23	0	3	0	0	0	.040	.042
71	892	76	4	2	14	2	0	0	2	0	7	0	0	0	0	0	.043	.044
72	958	12	3	1	18	0	0	0	3	0	5	0	1	0	0	0	.048	.049
73	962	13	2	1	15	0	0	0	3	0	4	0	0	0	0	0	.050	.051
74	965	14	1	0	13	0	0	0	3	0	2	0	0	0	0	0	.049	.049
75	968	12	1	0	14	0	0	0	3	0	2	0	0	0	0	0	.047	.048
76	969	13	1	0	12	0	0	0	2	0	2	0	0	0	0	0	.048	.048
77	962	12	1	0	19	0	0	0	3	0	2	0	1	0	0	0	.050	.050
78	950	24	1	0	19	0	0	0	3	0	2	0	1	0	0	0	.048	.048
79	963	19	1	0	12	0	0	0	2	0	2	0	0	0	0	0	.048	.049
80	966	12	1	0	15	0	0	0	3	0	2	0	0	0	0	0	.047	.047
81	877	23	2	0	84	1	0	0	5	0	4	0	3	0	0	0	.050	.051
82	921	53	2	2	13	1	0	0	3	0	4	0	0	0	0	0	.051	.052
83	967	13	1	0	13	0	0	0	3	0	2	0	0	0	0	0	.052	.052
84	970	12	1	0	12	0	0	0	2	0	2	0	0	0	0	0	.051	.052
85	955	11	2	0	24	0	0	0	3	0	3	0	1	0	0	0	.050	.050
86	951	28	1	0	15	0	0	0	3	0	1	0	0	0	0	0	.051	.051
87	957	17	1	0	19	0	0	0	3	0	2	0	1	0	0	0	.049	.049
88	959	21	1	0	14	0	0	0	3	0	2	0	0	0	0	0	.050	.051
89	936	18	1	0	35	1	0	0	4	0	3	0	1	0	0	0	.048	.048
90	947	35	1	0	11	0	0	0	2	0	2	0	0	0	0	0	.049	.049
91	971	11	1	0	12	0	0	0	2	0	2	0	0	0	0	0	.048	.048
92	965	12	1	0	16	0	0	0	3	0	2	0	0	0	0	0	.049	.049
93	957	18	1	0	18	0	0	0	3	0	2	0	1	0	0	0	.048	.048
94	922	26	1	0	41	1	0	0	4	0	3	0	1	0	0	0	.050	.050
95	928	51	1	0	15	1	0	0	3	0	2	0	0	0	0	0	.049	.049
96	965	13	1	1	15	0	0	0	3	0	2	0	0	0	0	0	.048	.048
97	968	15	1	0	11	0	0	0	2	0	1	0	0	0	0	0	.048	.048
98	970	11	1	0	13	0	0	0	3	0	1	0	0	0	0	0	.049	.049
99	968	13	1	0	13	0	0	0	3	0	1	0	0	0	0	0	.048	.048
100	969	12	1	0	13	0	0	0	3	0	2	0	0	0	0	0	.048	.047

APPENDIX N cont.

TABLE N.3

t	$p^{(11)}$	$p^{(21)}$	$p^{(31)}$	$p^{(41)}$	$p^{(12)}$	$p^{(22)}$	$p^{(32)}$	$p^{(42)}$	$p^{(13)}$	$p^{(23)}$	$p^{(33)}$	$p^{(43)}$	$p^{(14)}$	$p^{(24)}$	$p^{(34)}$	$p^{(44)}$	335 b_t	MSM b_t
1	876	92	0	3	21	2	0	0	3	0	2	0	1	0	0	0	.000	.000
2	958	15	1	1	18	0	0	0	3	0	4	0	1	0	0	0	.009	.009
3	963	13	1	1	15	0	0	0	3	0	4	0	0	0	0	0	.010	.010
4	965	12	1	0	15	0	0	0	3	0	4	0	0	0	0	0	.011	.011
5	960	13	1	0	18	0	0	0	3	0	4	0	1	0	0	0	.004	.004
6	964	13	1	1	14	0	0	0	3	0	4	0	0	0	0	0	.003	.003
7	948	13	1	0	27	0	0	0	3	0	6	0	1	0	0	0	-.005	-.005
8	958	17	1	1	14	0	0	0	3	0	6	0	0	0	0	0	-.007	-.007
9	929	20	1	0	38	1	0	0	4	0	6	0	1	0	0	0	-.001	-.001
10	952	23	1	1	14	1	0	0	3	0	5	0	0	0	0	0	.001	.001
11	967	13	1	0	12	0	0	0	2	0	4	0	0	0	0	0	.001	.001
12	970	11	0	0	12	0	0	0	2	0	3	0	0	0	0	0	.001	.002
13	961	12	1	0	18	0	0	0	3	0	4	0	1	0	0	0	-.001	-.001
14	954	13	1	1	21	0	0	0	3	0	6	0	1	0	0	0	-.003	-.003
15	873	38	0	0	73	2	1	0	5	0	5	0	2	0	0	0	.001	.001
16	902	78	0	1	12	1	0	0	2	0	3	0	0	0	0	0	-.000	-.000
17	965	13	0	0	14	0	0	0	3	0	3	0	0	0	0	0	.001	.001
18	917	20	1	0	50	1	0	0	4	0	6	0	2	0	0	0	-.002	-.00[illegible]
19	939	41	0	1	11	1	0	0	2	0	4	0	0	0	0	0	-.002	-.002
20	954	12	1	0	23	0	0	0	3	0	6	0	1	0	0	0	-.003	-.004
21	0	0	0	0	938	23	7	1	0	0	0	0	30	1	0	0	-.003	-.003
22	0	984	0	0	0	13	0	0	0	3	0	0	0	0	0	0	-.003	-.005
23	975	8	0	0	12	0	0	0	3	0	1	0	0	0	0	0	-.002	-.001
24	962	13	0	0	18	0	0	0	3	0	2	0	1	0	0	0	-.004	-.004
25	964	19	0	0	13	0	0	0	3	0	1	0	0	0	0	0	-.003	-.00[illegible]
26	971	12	0	0	12	0	0	0	2	0	2	0	0	0	0	0	-.003	-.001
27	969	12	0	0	13	0	0	0	3	0	2	0	0	0	0	0	-.002	-.002
28	970	13	0	0	12	0	0	0	2	0	2	0	0	0	0	0	-.003	-.003
29	972	11	0	0	11	0	0	0	2	0	2	0	0	0	0	0	-.003	-.003
30	945	12	1	0	32	0	0	0	4	0	4	0	1	0	0	0	-.001	-.001
31	953	25	1	1	13	0	0	0	3	0	4	0	0	0	0	0	-.000	-.000
32	965	13	0	0	14	0	0	0	3	0	3	0	0	0	0	0	-.001	-.00[illegible]
33	969	14	0	0	11	0	0	0	2	0	2	0	0	0	0	0	-.001	-.001
34	970	11	0	0	13	0	0	0	3	0	2	0	0	0	0	0	-.001	-.004
35	961	14	0	0	17	0	0	0	3	0	3	0	1	0	0	0	-.002	-.002
36	960	19	0	0	14	0	0	0	3	0	2	0	0	0	0	0	-.001	-.001
37	963	16	0	0	15	0	0	0	3	0	3	0	0	0	0	0	-.002	-.002
38	673	19	1	0	270	4	1	0	8	0	13	0	9	0	0	0	.002	.001
39	696	285	0	1	9	3	0	0	2	1	3	0	0	0	0	0	-.000	.000
40	556	27	4	1	353	4	2	0	8	0	33	0	11	0	0	0	.005	.005
41	0	1	0	0	564	363	32	9	0	0	0	0	18	12	1	0	.004	.004
42	6	460	0	260	0	251	0	9	0	4	1	1	0	8	0	0	-.006	.006
43	863	10	3	13	63	19	1	1	4	0	22	0	2	1	0	0	-.006	.00[illegible]
44	946	27	1	2	13	1	0	0	3	0	7	0	0	0	0	0	-.004	.003
45	967	10	1	0	13	0	0	0	3	0	5	0	0	0	0	0	-.006	.001
46	963	13	1	0	15	0	0	0	3	0	4	0	0	0	0	0	-.002	.004
47	957	16	1	0	18	0	0	0	3	0	4	0	1	0	0	0	-.008	-.003
48	961	17	1	0	14	0	0	0	3	0	4	0	0	0	0	0	-.005	-.002
49	966	12	1	0	14	0	0	0	3	0	4	0	0	0	0	0	-.003	-.000
50	940	18	1	0	31	0	0	0	4	0	5	0	1	0	0	0	-.009	-.00[illegible]

$p^{(ij)} \times 10^3$ and b_t (estimate of β_t) produced by 335 when applied to C8 Data with $\pi^{(33)} = .8$ and using the standard SSP

The last column is the MSM b_t for the same data

APPENDIX N cont.

TABLE N.3 cont.

51	953	26	0	1	12	0	0	0	2	0	3	0	0	0	0	0	-.009	-.007
52	932	13	1	1	42	1	0	0	4	0	6	0	1	0	0	0	-.003	-.002
53	902	66	0	0	22	1	0	0	3	0	3	0	1	0	0	0	-.007	-.006
54	962	17	1	1	13	0	0	0	3	0	4	0	0	0	0	0	-.008	-.007
55	968	12	0	0	13	0	0	0	3	0	3	0	0	0	0	0	-.009	-.008
56	937	16	1	0	36	0	0	0	4	0	5	0	1	0	0	0	-.006	-.005
57	950	24	1	1	15	1	0	0	3	0	5	0	0	0	0	0	-.005	-.004
58	928	12	2	1	38	1	0	0	4	0	12	0	1	0	0	0	-.002	-.001
59	928	47	0	0	16	1	0	0	3	0	4	0	0	0	0	0	-.004	-.003
60	964	13	1	1	14	0	0	0	3	0	4	0	0	0	0	0	-.005	-.004
61	945	18	0	0	28	0	0	0	4	0	3	0	1	0	0	0	-.003	-.002
62	942	18	1	2	24	1	0	0	3	0	7	0	1	0	0	0	-.001	.000
63	940	17	2	1	24	1	0	0	3	0	11	0	1	0	0	0	.001	.002
64	56	1	24	3	655	17	10	1	3	0	208	0	21	1	0	0	.021	.006
65	187	15	75	110	6	47	15	2	1	0	539	0	0	1	0	0	.046	.033
66	416	25	68	2	8	1	11	0	1	0	466	0	0	0	0	0	.050	.046
67	467	16	59	1	14	1	12	0	2	0	428	0	0	0	0	0	.073	.063
68	567	48	33	1	24	1	19	0	2	0	303	0	1	0	1	0	.037	.048
69	713	49	23	1	17	1	8	0	2	0	185	0	1	0	0	0	.057	.057
70	612	74	21	1	64	2	15	0	3	0	205	0	2	0	0	0	.028	.042
71	798	72	11	2	14	2	4	0	2	0	94	0	0	0	0	0	.042	.044
72	885	12	10	1	18	0	2	0	3	0	69	0	1	0	0	0	.049	.048
73	912	13	7	1	15	0	1	0	3	0	48	0	0	0	0	0	.052	.051
74	941	14	3	0	13	0	1	0	3	0	24	0	0	0	0	0	.049	.049
75	953	11	2	0	14	0	0	0	3	0	15	0	0	0	0	0	.047	.047
76	961	13	1	0	12	0	0	0	2	0	9	0	0	0	0	0	.048	.048
77	956	12	1	0	19	0	0	0	3	0	8	0	1	0	0	0	.051	.050
78	945	24	1	0	20	0	0	0	3	0	5	0	1	0	0	0	.048	.048
79	960	19	1	0	12	0	0	0	2	0	4	0	0	0	0	0	.048	.049
80	964	13	1	0	15	0	0	0	3	0	4	0	0	0	0	0	.047	.047
81	874	24	1	0	83	1	1	0	5	0	7	0	3	0	0	0	.051	.051
82	921	51	1	2	13	1	0	0	3	0	6	0	0	0	0	0	.052	.052
83	966	12	1	0	13	0	0	0	3	0	4	0	0	0	0	0	.052	.052
84	969	12	0	0	12	0	0	0	2	0	3	0	0	0	0	0	.052	.052
85	952	11	1	0	25	0	0	0	3	0	5	0	1	0	0	0	.050	.050
86	949	29	0	0	14	0	0	0	3	0	3	0	0	0	0	0	.051	.051
87	956	17	0	0	19	0	0	0	3	0	3	0	1	0	0	0	.050	.049
88	958	21	0	0	14	0	0	0	3	0	3	0	0	0	0	0	.051	.051
89	934	18	1	0	37	1	0	0	4	0	4	0	1	0	0	0	.049	.048
90	945	37	0	1	11	0	0	0	2	0	3	0	0	0	0	0	.049	.049
91	970	11	0	0	12	0	0	0	2	0	3	0	0	0	0	0	.049	.048
92	965	12	0	0	15	0	0	0	3	0	3	0	0	0	0	0	.050	.049
93	956	18	0	0	19	0	0	0	3	0	3	0	1	0	0	0	.048	.048
94	922	26	0	0	40	1	0	0	4	0	4	0	1	0	0	0	.050	.050
95	930	48	0	0	15	1	0	0	3	0	3	0	0	0	0	0	.049	.049
96	964	13	0	1	15	0	0	0	3	0	3	0	0	0	0	0	.048	.048
97	968	15	0	0	11	0	0	0	2	0	2	0	0	0	0	0	.048	.048
98	970	11	0	0	13	0	0	0	3	0	2	0	0	0	0	0	.049	.049
99	968	13	0	0	13	0	0	0	3	0	2	0	0	0	0	0	.048	.048
100	968	12	0	0	13	0	0	0	3	0	3	0	0	0	0	0	.048	.047

APPENDIX P

TABLE P.1

t	REALISATION 1	2	3	4	5	6	7	8	9	10
1	.004	-.003	-.001	.006	.000	.002	-.002	-.004	-.003	-.006
2	.000	.003	-.008	-.015	.009	.012	.002	-.001	-.005	-.001
3	.001	.012	.005	.012	.010	-.006	-.001	.020	.009	-.007
4	-.006	.008	.009	.008	.011	-.015	-.000	.017	.006	.002
5	.010	-.007	.014	-.001	.004	-.020	.004	.008	.000	.003
6	.007	-.003	.013	-.011	.003	-.017	.012	.001	.004	.008
7	.007	-.009	.014	-.008	-.005	-.014	.007	.004	-.006	.010
8	.005	-.008	.007	-.016	-.007	-.007	.002	.009	-.008	.004
9	.006	-.004	.008	-.009	-.001	-.000	.007	.010	-.006	.003
10	.000	-.001	.008	-.006	.001	-.003	.005	.004	-.006	.006
11	-.002	-.004	.005	-.007	.001	.005	.006	.005	-.001	.002
12	-.001	-.002	.006	-.004	.002	-.001	.001	.002	-.005	-.001
13	-.003	-.005	.004	-.005	-.001	-.004	.003	.002	-.001	-.002
14	-.001	-.005	.001	-.001	-.003	-.005	-.001	.003	.001	-.000
15	-.005	-.005	.002	-.005	.001	-.006	-.002	.004	-.003	.001
16	-.007	-.002	.002	-.006	-.000	-.006	-.004	.005	-.002	.003
17	-.009	-.001	.002	-.003	.001	-.005	-.003	.006	-.003	.004
18	-.008	-.000	.001	-.002	-.002	-.006	-.002	.004	.001	.004
19	-.005	-.000	-.001	-.002	-.002	-.006	-.000	.001	.002	.002
20	-.004	-.002	.001	-.001	-.004	-.007	-.002	.004	-.000	.002
21	-.004	-.001	-.000	-.000	-.001	-.005	-.003	.004	.002	.000
22	-.003	.000	.001	.003	.000	-.003	.001	.005	.001	.000
23	.002	.002	.002	.005	.002	-.001	.000	.008	.004	.002
24	.027	.008	.005	.008	.002	.001	.003	.010	.006	.003
25	.001	.015	.005	.011	.005	.003	.005	.016	.010	.004
26	.012	.030	.009	.012	.006	.001	.005	.013	.009	.006
27	.024	.024	.014	.013	.010	.006	.006	.012	.014	.008
28	.036	.020	.021	.014	.012	.009	.011	.012	.015	.011
29	.033	.026	.021	.015	.014	.022	.017	.010	.020	.013
30	.042	.026	.024	.019	.019	.032	.022	.015	.027	.013
31	.038	.023	.024	.019	.024	.035	.021	.018	.028	.013
32	.040	.022	.024	.022	.022	.043	.021	.021	.026	.014
33	.036	.023	.028	.020	.024	.037	.023	.022	.029	.016
34	.033	.024	.022	.018	.026	.038	.020	.024	.029	.017
35	.035	.025	.024	.021	.024	.031	.022	.024	.027	.016
36	.034	.027	.023	.024	.026	.029	.026	.028	.028	.015
37	.033	.024	.022	.023	.024	.030	.028	.027	.027	.014
38	.032	.022	.022	.025	.029	.026	.026	.024	.029	.018
39	.031	.024	.022	.027	.027	.021	.023	.023	.030	.018
40	.029	.026	.023	.033	.033	.026	.025	.024	.031	.021
41	.029	.026	.021	.032	.028	.027	.023	.023	.031	.020
42	.029	.024	.024	.026	.031	.028	.023	.023	.028	.021
43	.030	.024	.023	.024	.027	.033	.022	.023	.027	.024
44	.030	.028	.023	.021	.027	.031	.020	.025	.023	.020
45	.029	.027	.024	.026	.025	.033	.018	.024	.023	.021
46	.028	.027	.025	.028	.026	.028	.022	.025	.018	.019
47	.028	.028	.025	.029	.024	.031	.020	.027	.025	.020
48	.027	.028	.023	.030	.024	.031	.020	.027	.024	.023
49	.027	.029	.026	.026	.024	.030	.021	.028	.021	.034
50	.029	.028	.025	.025	.022	.030	.020	.031	.021	.023

b_t (estimate of β_t) produced by 335 with $\pi^{(33)} = \cdot 2$ to a growth change of size ½σ at time $t = 21$

i.e. $\beta_t = \begin{cases} 0 & t < 21 \\ 0{\cdot}025 & t \geqslant 21 \end{cases}$

APPENDIX P cont.

TABLE P.2

t	REALISATION 1	2	3	4	5	6	7	8	9	10
1	.004	-.003	-.001	.006	.000	.002	-.002	-.004	-.003	-.006
2	.000	.003	-.008	-.015	.009	.012	.002	-.001	-.005	-.001
3	.001	.012	.005	.012	.010	-.006	-.001	.020	.009	-.007
4	-.006	.008	.009	.008	.011	-.015	-.000	.017	.006	.002
5	.010	-.007	.014	-.001	.004	-.020	.004	.008	.000	.003
6	.007	-.003	.013	-.011	.003	-.017	.012	.001	.004	.008
7	.007	-.009	.014	-.008	-.005	-.014	.007	.004	-.006	.010
8	.005	-.008	.007	-.016	-.007	-.007	.002	.009	-.008	.003
9	.006	-.004	.008	-.009	-.001	.000	.007	.010	-.006	.003
10	.000	-.001	.008	-.006	.001	-.003	.005	.004	-.006	.006
11	-.002	-.004	.005	-.007	.001	.005	.006	.005	-.001	.002
12	-.000	-.002	.006	-.004	.001	-.001	.001	.002	-.005	-.001
13	-.003	-.005	.004	-.005	-.001	-.004	.003	.002	-.001	-.002
14	-.001	-.005	.001	-.001	-.003	-.005	-.001	.003	.001	-.000
15	-.005	-.005	.002	-.005	.001	-.006	-.002	.004	-.003	.001
16	-.007	-.002	.002	-.005	-.000	-.006	-.004	.005	-.002	.003
17	-.009	-.001	.003	-.003	.001	-.005	-.003	.006	-.003	.004
18	-.008	-.000	.001	-.002	-.002	-.006	-.002	.004	.001	.004
19	-.005	-.000	-.001	-.002	-.002	-.006	-.000	.001	.002	.002
20	-.004	-.002	.001	-.001	-.004	-.007	-.002	.004	-.001	.002
21	-.004	-.001	-.000	-.000	-.001	-.005	-.003	.004	.002	.000
22	-.003	.000	.001	.003	-.000	-.003	.001	.005	.001	.000
23	.002	.002	.002	.005	.002	-.002	.000	.008	.004	.002
24	.036	.009	.005	.007	.002	.000	.002	.010	.006	.003
25	.002	.015	.005	.010	.005	.003	.005	.016	.010	.004
26	.013	.032	.009	.012	.006	.001	.005	.013	.009	.006
27	.025	.024	.014	.013	.009	.005	.006	.011	.014	.008
28	.037	.020	.021	.014	.011	.009	.011	.011	.015	.011
29	.033	.027	.021	.015	.014	.028	.017	.010	.020	.013
30	.042	.026	.024	.019	.020	.034	.022	.015	.027	.012
31	.038	.023	.024	.019	.024	.035	.021	.018	.028	.013
32	.040	.022	.025	.022	.022	.044	.021	.021	.026	.014
33	.035	.023	.028	.019	.024	.037	.023	.022	.029	.016
34	.033	.024	.022	.018	.027	.038	.020	.024	.029	.017
35	.035	.025	.024	.021	.024	.030	.022	.024	.027	.016
36	.034	.027	.023	.024	.026	.028	.026	.028	.028	.015
37	.033	.024	.022	.023	.024	.029	.028	.027	.027	.014
38	.032	.022	.022	.025	.029	.025	.026	.024	.029	.018
39	.031	.024	.022	.027	.028	.021	.023	.023	.030	.018
40	.029	.025	.023	.034	.033	.026	.025	.024	.031	.021
41	.029	.026	.021	.032	.028	.026	.023	.023	.031	.020
42	.029	.024	.024	.026	.031	.028	.023	.023	.028	.021
43	.030	.024	.023	.024	.027	.033	.022	.023	.027	.024
44	.030	.028	.023	.021	.027	.030	.020	.025	.023	.020
45	.029	.027	.024	.026	.025	.033	.018	.024	.024	.021
46	.028	.027	.025	.028	.026	.028	.022	.025	.018	.019
47	.029	.028	.025	.029	.024	.031	.020	.027	.025	.020
48	.027	.028	.023	.030	.024	.031	.020	.027	.024	.023
49	.027	.029	.026	.026	.024	.030	.021	.028	.021	.036
50	.029	.028	.025	.025	.022	.030	.020	.031	.021	.023

b_t (estimate of β_t) produced by 335 with $\pi^{(33)} = .4$ to a growth change of size $\frac{1}{2}\sigma$ at time $t = 21$

i.e. $\beta_t = \begin{cases} 0 & t < 21 \\ 0.025 & t \geq 21 \end{cases}$

TABLE P.3

t	1	2	3	4	5	6	7	8	9	10
	REALISATION									
1	.004	-.003	-.001	.006	.000	.002	-.002	-.004	-.003	-.006
2	.000	.003	-.008	-.015	.009	.012	.002	-.001	-.005	-.001
3	.001	.012	.005	.012	.010	-.006	-.001	.021	.009	-.007
4	-.006	.008	.009	.008	.011	-.015	-.000	.017	.006	.002
5	.010	-.007	.014	-.001	.004	-.020	.004	.008	.000	.003
6	.007	-.003	.013	-.011	.003	-.017	.012	.001	.004	.008
7	.007	-.009	.014	-.008	-.005	-.014	.007	.004	-.006	.010
8	.005	-.008	.007	-.016	-.007	-.007	.002	.009	-.008	.003
9	.006	-.004	.008	-.009	-.001	.000	.007	.010	-.006	.003
10	.000	-.001	.008	-.006	.001	-.003	.005	.004	-.006	.006
11	-.002	-.004	.005	-.007	.001	.005	.006	.005	-.001	.002
12	-.000	-.002	.006	-.004	.001	-.001	.001	.002	-.005	-.001
13	-.003	-.005	.004	-.005	-.001	-.004	.003	.002	-.001	-.002
14	-.001	-.005	.001	-.001	-.003	-.005	-.001	.003	.001	.000
15	-.005	-.005	.002	-.005	.001	-.006	-.002	.004	-.003	.001
16	-.007	-.002	.002	-.005	-.000	-.006	-.004	.005	-.002	.003
17	-.009	-.001	.003	-.003	.001	-.005	-.003	.006	-.002	.004
18	-.008	-.000	.001	-.002	-.002	-.006	-.002	.004	.001	.004
19	-.005	-.000	-.001	-.002	-.002	-.006	-.000	.001	.002	.002
20	-.004	-.002	.001	-.001	-.004	-.007	-.002	.004	-.001	.002
21	-.004	-.001	.000	-.000	-.001	-.005	-.003	.004	.002	.001
22	-.003	-.000	.001	.003	-.000	-.003	.001	.005	.001	.000
23	.002	.001	.002	.004	.002	-.002	.000	.008	.004	.002
24	.047	.010	.004	.007	.002	.000	.002	.009	.006	.003
25	.002	.015	.004	.010	.004	.002	.004	.016	.010	.004
26	.014	.034	.009	.011	.006	.001	.004	.012	.008	.006
27	.027	.022	.013	.012	.009	.005	.005	.011	.014	.007
28	.038	.018	.022	.014	.011	.009	.011	.011	.014	.010
29	.034	.027	.021	.015	.013	.039	.016	.010	.020	.012
30	.042	.026	.024	.018	.021	.035	.021	.015	.029	.012
31	.038	.022	.024	.018	.024	.034	.020	.018	.028	.013
32	.040	.021	.024	.021	.022	.044	.020	.021	.025	.014
33	.035	.023	.029	.019	.025	.036	.023	.022	.029	.016
34	.032	.024	.022	.018	.027	.037	.020	.024	.029	.017
35	.035	.025	.023	.021	.024	.028	.022	.024	.027	.016
36	.034	.027	.023	.024	.027	.026	.027	.028	.028	.015
37	.033	.024	.022	.023	.025	.028	.028	.027	.027	.014
38	.032	.022	.022	.025	.030	.024	.026	.024	.029	.018
39	.031	.024	.022	.027	.028	.019	.023	.023	.030	.018
40	.029	.025	.023	.035	.034	.024	.025	.024	.031	.021
41	.029	.025	.021	.032	.029	.025	.023	.023	.031	.020
42	.029	.024	.024	.025	.031	.027	.023	.023	.028	.021
43	.030	.024	.023	.024	.027	.033	.022	.023	.027	.023
44	.030	.028	.023	.021	.027	.029	.020	.025	.024	.020
45	.029	.026	.024	.027	.026	.032	.018	.024	.024	.021
46	.028	.027	.025	.028	.026	.027	.022	.025	.018	.019
47	.029	.027	.024	.029	.024	.031	.020	.027	.026	.020
48	.027	.027	.023	.030	.024	.031	.020	.026	.024	.023
49	.028	.028	.026	.026	.024	.030	.021	.027	.021	.039
50	.029	.028	.025	.025	.022	.030	.020	.031	.021	.023

b_t (estimate of β_t) produced by 335 with $\pi^{(33)} = .6$ to a growth change of size $\frac{1}{2}\sigma$ at time $t=21$

i.e. $\beta_t = \begin{cases} 0 & t < 21 \\ 0.025 & t \geq 21 \end{cases}$

APPENDIX P cont.

TABLE P.4

t	REALISATION 1	2	3	4	5	6	7	8	9	10
1	.004	-.003	-.001	.006	.000	.002	-.002	-.004	-.003	-.006
2	.000	.003	-.008	-.015	.009	.012	.002	-.001	-.005	-.001
3	.001	.012	.005	.012	.010	-.006	-.001	.021	.009	-.007
4	-.006	.008	.009	.007	.011	-.015	-.000	.017	.006	.002
5	.010	-.008	.014	-.001	.004	-.020	.004	.008	.000	.003
6	.007	-.003	.013	-.011	.003	-.017	.012	.001	.004	.008
7	.007	-.009	.013	-.007	-.005	-.014	.007	.004	-.006	.010
8	.005	-.008	.007	-.016	-.007	-.007	.002	.010	-.008	.003
9	.006	-.004	.008	-.009	-.001	.000	.007	.010	-.006	.003
10	-.000	-.001	.008	-.006	.001	-.003	.005	.004	-.006	.006
11	-.002	-.004	.005	-.007	.001	.005	.006	.005	-.001	.002
12	-.000	-.002	.006	-.004	.001	-.002	.000	.002	-.005	-.001
13	-.003	-.005	.004	-.005	-.001	-.005	.003	.002	-.001	-.001
14	-.001	-.005	.002	-.001	-.003	-.005	-.001	.003	.001	.000
15	-.004	-.005	.002	-.006	.001	-.006	-.002	.004	-.003	.001
16	-.007	-.002	.002	-.005	-.000	-.006	-.004	.005	-.002	.002
17	-.008	-.001	.003	-.003	.001	-.005	-.003	.006	-.002	.003
18	-.007	-.001	.001	-.003	-.002	-.006	-.002	.004	.001	.004
19	-.005	-.001	-.000	-.002	-.002	-.006	-.000	.002	.002	.002
20	-.004	-.003	.001	-.002	-.003	-.007	-.002	.004	-.001	.002
21	-.005	-.001	.000	-.001	-.001	-.005	-.003	.004	.001	.001
22	-.004	-.000	.001	.003	-.000	-.004	.001	.004	.000	.001
23	.002	.001	.002	.003	.001	-.002	-.000	.007	.003	.002
24	.060	.013	.004	.005	.001	-.001	.002	.008	.005	.003
25	.003	.013	.004	.008	.003	.001	.004	.018	.010	.003
26	.016	.041	.009	.009	.005	-.000	.003	.011	.007	.005
27	.030	.016	.014	.010	.008	.004	.004	.010	.013	.006
28	.042	.013	.024	.012	.009	.009	.012	.011	.013	.009
29	.035	.028	.018	.013	.012	.060	.017	.010	.021	.011
30	.044	.026	.022	.018	.028	.033	.020	.015	.036	.010
31	.039	.021	.023	.017	.028	.028	.018	.018	.028	.011
32	.041	.020	.024	.021	.020	.046	.019	.021	.024	.013
33	.035	.023	.029	.018	.025	.030	.022	.021	.029	.015
34	.032	.024	.020	.017	.029	.035	.019	.024	.029	.016
35	.035	.025	.023	.021	.024	.018	.022	.024	.027	.016
36	.034	.028	.023	.024	.027	.022	.028	.028	.028	.014
37	.033	.024	.021	.023	.024	.026	.029	.026	.026	.014
38	.032	.022	.022	.025	.032	.020	.026	.023	.029	.018
39	.031	.024	.021	.027	.029	.014	.023	.023	.030	.017
40	.029	.025	.023	.038	.036	.023	.026	.024	.032	.021
41	.029	.025	.021	.031	.029	.025	.023	.023	.032	.020
42	.029	.024	.024	.026	.032	.026	.023	.023	.028	.020
43	.030	.024	.023	.024	.027	.036	.022	.023	.027	.023
44	.030	.028	.023	.022	.027	.030	.020	.025	.023	.020
45	.029	.026	.024	.027	.025	.033	.018	.024	.025	.020
46	.028	.026	.025	.028	.026	.027	.022	.025	.019	.019
47	.029	.027	.024	.029	.024	.031	.020	.026	.027	.020
48	.027	.027	.023	.030	.024	.031	.020	.026	.025	.024
49	.028	.028	.025	.026	.024	.030	.021	.027	.022	.045
50	.029	.028	.024	.026	.022	.030	.020	.030	.022	.023

b_t (estimate of β_t) produced by 335 with $\pi^{(33)} = .8$ to a growth change of size ½ σ at time $t=21$

i.e. $$\beta_t = \begin{cases} 0 & t < 21 \\ 0.025 & t \geq 21 \end{cases}$$

TABLE P.5

t	REALISATION 1	2	3	4	5	6	7	8	9	10
1	.004	-.003	-.001	.006	.000	.002	-.002	-.004	-.003	-.006
2	.000	.003	-.008	-.015	.009	.012	.002	-.001	-.005	-.001
3	.001	.012	.005	.012	.010	-.006	-.001	.020	.009	-.007
4	-.006	.008	.009	.008	.011	-.015	-.000	.017	.006	.002
5	.010	-.007	.014	-.001	.004	-.020	.004	.008	.000	.003
6	.007	-.003	.013	-.011	.003	-.017	.012	.001	.004	.008
7	.007	-.009	.014	-.008	-.005	-.014	.007	.004	-.006	.010
8	.005	-.008	.007	-.016	-.007	-.007	.002	.009	-.008	.004
9	.006	-.004	.008	-.009	-.001	-.000	.007	.010	-.006	.003
10	.000	-.001	.008	-.006	.001	-.003	.005	.004	-.006	.006
11	-.002	-.004	.005	-.007	.001	.005	.006	.005	-.001	.002
12	-.001	-.002	.006	-.004	.002	-.001	.001	.002	-.005	-.001
13	-.003	-.005	.004	-.005	-.001	-.004	.003	.002	-.001	-.002
14	-.001	-.005	.001	-.001	-.003	-.005	-.001	.003	.001	-.000
15	-.005	-.005	.002	-.005	.001	-.006	-.002	.004	-.003	.001
16	-.007	-.002	.002	-.006	-.000	-.006	-.004	.005	-.002	.003
17	-.009	-.001	.002	-.003	.001	-.005	-.003	.006	-.003	.004
18	-.008	-.000	.001	-.002	-.002	-.006	-.002	.004	.001	.004
19	-.005	-.000	-.001	-.002	-.002	-.006	-.000	.001	.002	.002
20	-.004	-.002	.001	-.001	-.004	-.007	-.002	.004	-.000	.002
21	-.003	-.000	.001	.001	-.001	-.004	-.002	.005	.003	.001
22	-.001	.002	.003	.006	.003	-.001	.001	.007	.003	.002
23	.005	.007	.007	.020	.009	.007	.005	.014	.008	.006
24	.059	.027	.016	.030	.013	.020	.013	.027	.018	.011
25	.020	.049	.025	.038	.024	.030	.023	.048	.041	.019
26	.032	.072	.042	.040	.031	.026	.026	.040	.035	.030
27	.050	.051	.051	.040	.039	.036	.030	.036	.047	.038
28	.066	.042	.057	.042	.042	.041	.041	.036	.045	.045
29	.059	.052	.054	.042	.044	.051	.047	.034	.054	.047
30	.069	.050	.053	.047	.050	.053	.050	.041	.060	.044
31	.063	.045	.052	.046	.052	.054	.048	.045	.058	.043
32	.066	.044	.051	.049	.050	.059	.047	.048	.055	.043
33	.060	.046	.054	.046	.051	.057	.049	.049	.056	.045
34	.056	.048	.049	.044	.052	.058	.046	.051	.055	.045
35	.059	.049	.049	.047	.049	.053	.047	.051	.053	.044
36	.058	.051	.049	.049	.051	.053	.051	.054	.054	.041
37	.057	.048	.047	.048	.049	.053	.052	.053	.052	.040
38	.056	.046	.047	.050	.053	.050	.050	.049	.054	.044
39	.055	.048	.047	.052	.052	.047	.048	.048	.054	.044
40	.053	.050	.048	.057	.057	.051	.050	.050	.056	.046
41	.054	.050	.046	.056	.053	.051	.048	.049	.055	.046
42	.053	.049	.049	.051	.055	.053	.048	.048	.053	.046
43	.054	.049	.048	.048	.051	.058	.047	.047	.051	.048
44	.055	.052	.048	.046	.051	.055	.045	.050	.048	.045
45	.054	.052	.049	.051	.050	.058	.043	.049	.048	.045
46	.052	.052	.050	.053	.050	.053	.047	.050	.043	.044
47	.053	.053	.049	.054	.049	.056	.045	.052	.050	.044
48	.052	.053	.048	.055	.049	.056	.045	.052	.049	.048
49	.052	.054	.050	.051	.049	.055	.046	.053	.046	.059
50	.053	.053	.050	.050	.046	.055	.045	.056	.046	.048

b_t (estimate of β_t) produced by 335 with $\pi^{(33)} = \cdot 2$ to a growth change of size 1σ at time $t = 21$

i.e. $\beta_t = \begin{cases} 0 & t < 21 \\ 0.050 & t \geq 21 \end{cases}$

APPENDIX P cont.

TABLE P.6

t	REALISATION 1	2	3	4	5	6	7	8	9	10
1	.004	-.003	-.001	.006	.000	.002	-.002	-.004	-.003	-.006
2	.000	.003	-.008	-.015	.009	.012	.002	-.001	-.005	-.001
3	.001	.012	.005	.012	.010	-.006	-.001	.020	.009	-.007
4	-.006	.008	.009	.008	.011	-.015	-.000	.017	.006	.002
5	.010	-.007	.014	-.001	.004	-.020	.004	.008	.000	.003
6	.007	-.003	.013	-.011	.003	-.017	.012	.001	.004	.008
7	.007	-.009	.014	-.008	-.005	-.014	.007	.004	-.006	.010
8	.005	-.008	.007	-.016	-.007	-.007	.002	.009	-.008	.003
9	.006	-.004	.008	-.009	-.001	.000	.007	.010	-.006	.003
10	.000	-.001	.008	-.006	.001	-.003	.005	.004	-.006	.006
11	-.002	-.004	.005	-.007	.001	.005	.006	.005	-.001	.002
12	-.000	-.002	.006	-.004	.001	-.001	.001	.002	-.005	-.001
13	-.003	-.005	.004	-.005	-.001	-.004	.003	.002	-.001	-.002
14	-.001	-.005	.001	-.001	-.003	-.005	-.001	.003	.001	-.000
15	-.005	-.005	.002	-.005	.001	-.006	-.002	.004	-.003	.001
16	-.007	-.002	.002	-.005	-.000	-.006	-.004	.005	-.002	.003
17	-.009	-.001	.003	-.003	.001	-.005	-.003	.006	-.003	.004
18	-.008	-.000	.001	-.002	-.002	-.006	-.002	.004	.001	.004
19	-.005	-.000	-.001	-.002	-.002	-.006	-.000	.001	.002	.002
20	-.004	-.002	.001	-.001	-.004	-.007	-.002	.004	-.001	.002
21	-.004	-.001	.001	.000	-.001	-.004	-.002	.005	.003	.001
22	-.001	.002	.003	.007	.002	-.001	.001	.007	.003	.002
23	.007	.007	.007	.020	.009	.008	.005	.015	.008	.006
24	.075	.042	.018	.031	.013	.022	.013	.028	.020	.011
25	.026	.052	.026	.039	.025	.032	.025	.055	.046	.020
26	.037	.073	.046	.041	.032	.026	.027	.040	.035	.032
27	.053	.051	.053	.040	.041	.037	.031	.035	.048	.040
28	.067	.041	.058	.042	.043	.042	.044	.035	.045	.047
29	.059	.051	.054	.042	.045	.053	.049	.033	.055	.048
30	.069	.050	.053	.047	.051	.054	.051	.041	.060	.045
31	.063	.045	.052	.046	.053	.054	.048	.045	.058	.043
32	.065	.044	.051	.049	.050	.059	.047	.049	.054	.044
33	.060	.046	.054	.046	.051	.057	.049	.049	.056	.045
34	.056	.048	.048	.044	.052	.058	.046	.051	.055	.045
35	.059	.049	.049	.047	.050	.053	.047	.051	.053	.044
36	.058	.051	.049	.049	.051	.052	.051	.054	.053	.041
37	.057	.048	.047	.048	.049	.053	.052	.053	.052	.040
38	.056	.046	.047	.050	.053	.050	.050	.049	.053	.044
39	.055	.048	.047	.052	.052	.047	.048	.048	.054	.044
40	.053	.050	.048	.058	.057	.051	.050	.050	.055	.046
41	.054	.050	.046	.056	.052	.051	.048	.049	.055	.045
42	.053	.049	.049	.050	.054	.052	.048	.048	.053	.046
43	.054	.049	.048	.048	.051	.058	.047	.048	.051	.048
44	.055	.052	.048	.046	.051	.055	.045	.050	.048	.045
45	.054	.051	.049	.051	.050	.057	.043	.049	.048	.045
46	.053	.051	.050	.053	.050	.053	.047	.050	.043	.044
47	.053	.053	.049	.054	.049	.056	.045	.052	.050	.044
48	.052	.052	.048	.055	.049	.056	.045	.052	.049	.048
49	.052	.054	.050	.051	.049	.055	.046	.052	.046	.061
50	.053	.053	.049	.050	.047	.055	.045	.056	.046	.048

b_t (estimate of β_t) produced by 335 with $\pi^{(33)} = .4$ to a growth change of size 1σ at time $t = 21$

i.e. $\beta_t = \begin{cases} 0 & t < 21 \\ 0.050 & t \geq 21 \end{cases}$

TABLE P.7

t	REALISATION 1	2	3	4	5	6	7	8	9	10
1	.004	-.003	-.001	.006	.000	.002	-.002	-.004	-.003	-.006
2	.000	.003	-.008	-.015	.009	.012	.002	-.001	-.005	-.001
3	.001	.012	.005	.012	.010	-.006	-.001	.021	.009	-.007
4	-.006	.008	.009	.008	.011	-.015	-.000	.017	.006	.002
5	.010	-.007	.014	-.001	.004	-.020	.004	.008	.000	.003
6	.007	-.003	.013	-.011	.003	-.017	.012	.001	.004	.008
7	.007	-.009	.014	-.008	-.005	-.014	.007	.004	-.006	.010
8	.005	-.008	.007	-.016	-.007	-.007	.002	.009	-.008	.003
9	.006	-.004	.008	-.009	-.001	.000	.007	.010	-.006	.003
10	.000	-.001	.008	-.006	.001	-.003	.005	.004	-.006	.006
11	-.002	-.004	.005	-.007	.001	.005	.006	.005	-.001	.002
12	-.000	-.002	.006	-.004	.001	-.001	.001	.002	-.005	-.001
13	-.003	-.005	.004	-.005	-.001	-.004	.003	.002	-.001	-.002
14	-.001	-.005	.001	-.001	-.003	-.005	-.001	.003	.001	.000
15	-.005	-.005	.002	-.005	.001	-.006	-.002	.004	-.003	.001
16	-.007	-.002	.002	-.005	-.000	-.006	-.004	.005	-.002	.003
17	-.009	-.001	.003	-.003	.001	-.005	-.003	.006	-.002	.004
18	-.008	-.000	.001	-.002	-.002	-.006	-.002	.004	.001	.004
19	-.005	-.000	-.001	-.002	-.002	-.006	-.000	.001	.002	.002
20	-.004	-.002	.001	-.001	-.004	-.007	-.002	.004	-.001	.002
21	-.004	-.001	.001	.000	-.001	-.004	-.002	.004	.003	.001
22	-.001	.002	.003	.008	.002	-.001	.002	.007	.003	.002
23	.010	.008	.006	.021	.009	.008	.004	.016	.008	.006
24	.090	.058	.021	.031	.013	.024	.013	.029	.023	.011
25	.030	.053	.026	.040	.027	.033	.028	.064	.052	.020
26	.039	.074	.052	.040	.034	.024	.027	.038	.033	.034
27	.056	.045	.055	.040	.044	.038	.032	.032	.048	.043
28	.069	.035	.060	.042	.044	.044	.049	.034	.044	.049
29	.060	.052	.054	.042	.046	.058	.051	.031	.055	.049
30	.070	.050	.053	.048	.053	.056	.052	.042	.061	.044
31	.063	.044	.051	.047	.054	.056	.049	.047	.059	.043
32	.066	.042	.050	.049	.051	.061	.047	.050	.054	.043
33	.060	.045	.054	.046	.051	.058	.049	.050	.055	.045
34	.056	.047	.047	.044	.052	.059	.046	.052	.055	.045
35	.059	.049	.048	.047	.050	.053	.047	.052	.053	.043
36	.058	.051	.048	.049	.051	.052	.051	.055	.053	.041
37	.057	.048	.046	.048	.049	.053	.052	.053	.051	.039
38	.056	.046	.046	.050	.054	.050	.050	.050	.053	.044
39	.055	.048	.046	.052	.052	.046	.048	.048	.054	.043
40	.053	.050	.047	.058	.057	.051	.050	.050	.055	.046
41	.054	.050	.045	.056	.052	.051	.048	.049	.055	.045
42	.053	.049	.049	.050	.055	.052	.048	.049	.053	.046
43	.054	.049	.048	.048	.051	.057	.047	.048	.051	.048
44	.055	.052	.047	.046	.051	.055	.045	.050	.048	.045
45	.054	.051	.048	.051	.050	.057	.043	.049	.049	.045
46	.053	.051	.050	.053	.050	.053	.047	.050	.043	.044
47	.053	.052	.049	.054	.049	.056	.045	.052	.051	.044
48	.052	.052	.048	.055	.049	.056	.045	.051	.049	.048
49	.052	.053	.050	.050	.049	.055	.046	.052	.046	.064
50	.054	.053	.049	.050	.047	.055	.045	.056	.046	.047

b_t (estimate of β_t) produced by 335 with $\pi^{(33)} = .6$ to a growth change of size 1σ at time $t=21$

i.e. $\beta_t = \begin{cases} 0 & t < 21 \\ 0.050 & t \geq 21 \end{cases}$

TABLE P.8

t	REALISATION 1	2	3	4	5	6	7	8	9	10
1	.004	-.003	-.001	.006	.000	.002	-.002	-.004	-.003	-.006
2	.000	.003	-.008	-.015	.009	.012	.002	-.001	-.005	-.001
3	.001	.012	.005	.012	.010	-.006	-.001	.021	.009	-.007
4	-.006	.008	.009	.007	.011	-.015	-.000	.017	.006	.002
5	.010	-.008	.014	-.001	.004	-.020	.004	.008	.000	.003
6	.007	-.003	.013	-.011	.003	-.017	.012	.001	.004	.008
7	.007	-.009	.013	-.007	-.005	-.014	.007	.004	-.006	.010
8	.005	-.008	.007	-.016	-.007	-.007	.002	.010	-.008	.003
9	.006	-.004	.008	-.009	-.001	.000	.007	.010	-.006	.003
10	-.000	-.001	.008	-.006	.001	-.003	.005	.004	-.006	.006
11	-.002	-.004	.005	-.007	.001	.005	.006	.005	-.001	.002
12	-.000	-.002	.006	-.004	.001	-.002	.000	.002	-.005	-.001
13	-.003	-.005	.004	-.005	-.001	-.005	.003	.002	-.001	-.001
14	-.001	-.005	.002	-.001	-.003	-.005	-.001	.003	.001	.000
15	-.004	-.005	.002	-.006	.001	-.006	-.002	.004	-.003	.001
16	-.007	-.002	.002	-.005	-.000	-.006	-.004	.005	-.002	.002
17	-.008	-.001	.003	-.003	.001	-.005	-.003	.006	-.002	.003
18	-.007	-.001	.001	-.003	-.002	-.006	-.002	.004	.001	.004
19	-.005	-.001	-.000	-.002	-.002	-.006	-.000	.002	.002	.002
20	-.004	-.003	.001	-.002	-.003	-.007	-.002	.004	-.001	.002
21	-.004	-.001	.001	-.000	-.000	-.004	-.002	.004	.002	.001
22	-.002	.002	.003	.010	.002	-.001	.002	.006	.002	.002
23	.015	.008	.006	.022	.010	.010	.003	.018	.008	.006
24	.103	.075	.026	.030	.011	.028	.014	.031	.029	.010
25	.031	.050	.024	.041	.035	.035	.036	.079	.060	.021
26	.041	.078	.064	.039	.039	.016	.025	.029	.025	.041
27	.061	.031	.060	.037	.052	.044	.032	.021	.050	.047
28	.075	.022	.062	.041	.046	.050	.067	.030	.041	.054
29	.061	.057	.048	.042	.047	.079	.056	.028	.061	.048
30	.073	.050	.050	.050	.060	.059	.054	.049	.067	.038
31	.064	.040	.049	.047	.057	.056	.047	.053	.059	.039
32	.066	.040	.048	.051	.050	.065	.046	.055	.051	.042
33	.059	.046	.055	.046	.051	.057	.049	.053	.055	.045
34	.056	.048	.042	.043	.053	.059	.044	.055	.054	.044
35	.059	.049	.046	.047	.049	.051	.047	.053	.051	.042
36	.058	.053	.046	.050	.051	.051	.052	.058	.052	.039
37	.057	.048	.044	.048	.048	.052	.053	.055	.050	.038
38	.056	.045	.045	.050	.054	.048	.050	.049	.053	.044
39	.055	.048	.045	.052	.052	.044	.047	.048	.054	.043
40	.053	.050	.047	.060	.058	.049	.050	.050	.055	.047
41	.054	.050	.044	.056	.053	.050	.047	.049	.055	.045
42	.053	.049	.049	.051	.055	.051	.047	.048	.052	.046
43	.054	.049	.047	.050	.051	.058	.046	.047	.051	.049
44	.054	.053	.047	.048	.051	.054	.044	.050	.048	.045
45	.054	.051	.048	.052	.050	.057	.043	.049	.049	.045
46	.053	.051	.050	.053	.050	.052	.047	.050	.044	.044
47	.053	.052	.049	.053	.049	.055	.045	.052	.051	.044
48	.052	.052	.048	.054	.049	.055	.045	.051	.049	.048
49	.053	.053	.050	.050	.049	.054	.045	.052	.047	.068
50	.053	.053	.049	.050	.047	.054	.044	.055	.047	.048

b_t (estimate of β_t) produced by 33S with $\pi^{(33)} = .8$ to a growth change of size 1σ at time $t = 21$

i.e. $\beta_t = \begin{cases} 0 & t < 21 \\ 0.050 & t \geq 21 \end{cases}$

APPENDIX P cont.

TABLE P.9

t	REALISATION 1	2	3	4	5	6	7	8	9	10
1	.004	-.003	-.001	.006	.000	.002	-.002	-.004	-.003	-.006
2	.000	.003	-.008	-.015	.009	.012	.002	-.001	-.005	-.001
3	.001	.012	.005	.012	.010	-.006	-.001	.020	.009	-.007
4	-.006	.008	.009	.008	.011	-.015	-.000	.017	.006	.002
5	.010	-.007	.014	-.001	.004	-.020	.004	.008	.000	.003
6	.007	-.003	.013	-.011	.003	-.017	.012	.001	.004	.008
7	.007	-.009	.014	-.008	-.005	-.014	.007	.004	-.006	.010
8	.005	-.008	.007	-.016	-.007	-.007	.002	.009	-.008	.004
9	.006	-.004	.008	-.009	-.001	-.000	.007	.010	-.006	.003
10	.000	-.001	.008	-.006	.001	-.003	.005	.004	-.006	.006
11	-.002	-.004	.005	-.007	.001	.005	.006	.005	-.001	.002
12	-.001	-.002	.006	-.004	.002	-.001	.001	.002	-.005	-.001
13	-.003	-.005	.004	-.005	-.001	-.004	.003	.002	-.001	-.002
14	-.001	-.005	.001	-.001	-.003	-.005	-.001	.003	.001	-.000
15	-.005	-.005	.002	-.005	.001	-.006	-.002	.004	-.003	.001
16	-.007	-.002	.002	-.006	-.000	-.006	-.004	.005	-.002	.003
17	-.009	-.001	.002	-.003	.001	-.005	-.003	.006	-.003	.004
18	-.008	-.000	.001	-.002	-.002	-.006	-.002	.004	.001	.004
19	-.005	-.000	-.001	-.002	-.002	-.006	-.000	.001	.002	.002
20	-.004	-.002	.001	-.001	-.004	-.007	-.002	.004	-.000	.002
21	-.003	.000	.001	.001	-.000	-.004	-.002	.006	.004	.002
22	.002	.005	.005	.009	.008	.005	.000	.010	.005	.004
23	.011	.027	.019	.045	.034	.038	.014	.025	.010	.010
24	.103	.084	.059	.059	.043	.057	.030	.065	.045	.043
25	.055	.081	.059	.068	.060	.063	.045	.095	.076	.058
26	.062	.095	.079	.067	.064	.052	.044	.070	.057	.069
27	.078	.079	.082	.065	.070	.063	.050	.058	.073	.073
28	.091	.070	.085	.066	.070	.068	.067	.057	.068	.077
29	.083	.077	.080	.067	.071	.078	.072	.054	.080	.076
30	.094	.076	.079	.073	.076	.078	.075	.064	.086	.070
31	.087	.071	.077	.072	.077	.079	.072	.069	.084	.068
32	.090	.070	.076	.075	.075	.083	.071	.073	.079	.068
33	.084	.072	.079	.071	.075	.081	.072	.073	.080	.069
34	.081	.073	.073	.068	.076	.082	.070	.076	.079	.069
35	.083	.074	.074	.071	.074	.078	.071	.075	.077	.068
36	.082	.076	.073	.074	.075	.077	.075	.079	.078	.065
37	.082	.073	.072	.073	.074	.078	.076	.077	.076	.064
38	.081	.071	.072	.075	.077	.075	.074	.074	.078	.068
39	.080	.073	.071	.077	.076	.072	.072	.073	.079	.068
40	.078	.075	.073	.082	.081	.076	.074	.074	.080	.071
41	.078	.075	.071	.081	.077	.076	.072	.073	.080	.070
42	.078	.074	.074	.076	.079	.077	.072	.073	.078	.070
43	.079	.074	.073	.073	.076	.082	.071	.072	.076	.073
44	.079	.077	.072	.071	.076	.080	.069	.075	.073	.070
45	.079	.076	.073	.076	.075	.083	.067	.074	.073	.070
46	.077	.077	.075	.077	.075	.078	.071	.075	.068	.069
47	.078	.078	.074	.078	.073	.081	.070	.077	.075	.069
48	.077	.078	.073	.080	.074	.081	.070	.077	.074	.073
49	.077	.079	.075	.075	.074	.080	.070	.077	.071	.084
50	.078	.078	.074	.075	.071	.080	.070	.081	.071	.073

b_t (estimate of β_t) produced by 335 with $\pi^{(33)} = .2$

to a growth change of size 1.5σ at time $t=21$

i.e. $\beta_t = \begin{cases} 0 & t < 21 \\ 0.075 & t \geq 21 \end{cases}$

APPENDIX P cont.

TABLE P.10

t	REALISATION 1	2	3	4	5	6	7	8	9	10
1	.004	-.003	-.001	.006	.000	.002	-.002	-.004	-.003	-.006
2	.000	.003	-.008	-.015	.009	.012	.002	-.001	-.005	-.001
3	.001	.012	.005	.012	.010	-.006	-.001	.020	.009	-.007
4	-.006	.008	.009	.008	.011	-.015	-.000	.017	.006	.002
5	.010	-.007	.014	-.001	.004	-.020	.004	.008	.000	.003
6	.007	-.003	.013	-.011	.003	-.017	.012	.001	.004	.008
7	.007	-.009	.014	-.008	-.005	-.014	.007	.004	-.006	.010
8	.005	-.008	.007	-.016	-.007	-.007	.002	.009	-.008	.003
9	.006	-.004	.008	-.009	-.001	.000	.007	.010	-.006	.003
10	.000	-.001	.008	-.006	.001	-.003	.005	.004	-.006	.006
11	-.002	-.004	.005	-.007	.001	.005	.006	.005	-.001	.002
12	-.000	-.002	.006	-.004	.001	-.001	.001	.002	-.005	-.001
13	-.003	-.005	.004	-.005	-.001	-.004	.003	.002	-.001	-.002
14	-.001	-.005	.001	-.001	-.003	-.005	-.001	.003	.001	-.000
15	-.005	-.005	.002	-.005	.001	-.006	-.002	.004	-.003	.001
16	-.007	-.002	.002	-.005	-.000	-.006	-.004	.005	-.002	.003
17	-.009	-.001	.003	-.003	.001	-.005	-.003	.006	-.003	.004
18	-.008	-.000	.001	-.002	-.002	-.006	-.002	.004	.001	.004
19	-.005	-.000	-.001	-.002	-.002	-.006	-.000	.001	.002	.002
20	-.004	-.002	.001	-.001	-.004	-.007	-.002	.004	-.001	.002
21	-.003	.000	.001	.001	-.000	-.004	-.002	.005	.004	.002
22	.002	.006	.005	.011	.009	.007	.000	.010	.005	.004
23	.018	.034	.022	.049	.038	.043	.014	.032	.012	.011
24	.120	.099	.067	.060	.044	.060	.032	.069	.056	.046
25	.067	.083	.060	.068	.062	.064	.047	.098	.083	.060
26	.069	.096	.081	.067	.065	.052	.045	.070	.059	.070
27	.082	.078	.082	.064	.071	.064	.051	.058	.073	.074
28	.093	.069	.085	.066	.071	.069	.069	.057	.068	.077
29	.084	.077	.080	.067	.071	.079	.073	.054	.079	.075
30	.094	.075	.078	.073	.077	.079	.075	.064	.085	.070
31	.088	.071	.077	.071	.078	.079	.072	.069	.083	.068
32	.090	.069	.076	.074	.075	.083	.071	.073	.079	.068
33	.084	.071	.078	.071	.075	.081	.073	.073	.080	.069
34	.081	.073	.073	.068	.076	.082	.070	.076	.079	.069
35	.083	.073	.074	.071	.074	.078	.071	.075	.077	.067
36	.082	.076	.073	.074	.075	.077	.075	.079	.078	.065
37	.082	.073	.071	.073	.074	.078	.076	.077	.076	.064
38	.081	.071	.072	.075	.078	.075	.074	.074	.078	.068
39	.080	.073	.071	.077	.076	.072	.072	.073	.079	.068
40	.078	.075	.072	.083	.081	.076	.074	.074	.080	.071
41	.078	.075	.071	.081	.077	.076	.072	.073	.080	.070
42	.078	.074	.074	.075	.079	.077	.072	.073	.078	.070
43	.079	.073	.073	.073	.076	.082	.071	.072	.076	.073
44	.079	.077	.072	.071	.076	.080	.069	.075	.073	.070
45	.079	.076	.073	.076	.075	.082	.067	.074	.073	.070
46	.077	.076	.075	.077	.075	.078	.071	.075	.068	.069
47	.078	.077	.074	.078	.073	.081	.070	.076	.075	.069
48	.077	.077	.073	.080	.074	.081	.070	.076	.074	.073
49	.077	.079	.075	.075	.074	.080	.070	.077	.071	.086
50	.078	.078	.074	.075	.071	.080	.070	.081	.071	.073

b_t (estimates of β_t) produced by 335 with $\pi^{(33)} = .4$ to a growth change of size 1.5σ at time $t=21$

i.e. $\beta_t = \begin{cases} 0 & t < 21 \\ 0.075 & t \geq 21 \end{cases}$

TABLE P.11

t	REALISATION 1	2	3	4	5	6	7	8	9	10
1	.004	-.003	-.001	.006	.000	.002	-.002	-.004	-.003	-.006
2	.000	.003	-.008	-.015	.009	.012	.002	-.001	-.005	-.001
3	.001	.012	.005	.012	.010	-.006	-.001	.021	.009	-.007
4	-.006	.008	.009	.008	.011	-.015	-.000	.017	.006	.002
5	.010	-.007	.014	-.001	.004	-.020	.004	.008	.000	.003
6	.007	-.003	.013	-.011	.003	-.017	.012	.001	.004	.008
7	.007	-.009	.014	-.008	-.005	-.014	.007	.004	-.006	.010
8	.005	-.008	.007	-.016	-.007	-.007	.002	.009	-.008	.003
9	.006	-.004	.008	-.009	-.001	.000	.007	.010	-.006	.003
10	.000	-.001	.008	-.006	.001	-.003	.005	.004	-.006	.006
11	-.002	-.004	.005	-.007	.001	.005	.006	.005	-.001	.002
12	-.000	-.002	.006	-.004	.001	-.001	.001	.002	-.005	-.001
13	-.003	-.005	.004	-.005	-.001	-.004	.003	.002	-.001	-.002
14	-.001	-.005	.001	-.001	-.003	-.005	-.001	.003	.001	.000
15	-.005	-.005	.002	-.005	.001	-.006	-.002	.004	-.003	.001
16	-.007	-.002	.002	-.005	-.000	-.006	-.004	.005	-.002	.003
17	-.009	-.001	.003	-.003	.001	-.005	-.003	.006	-.002	.004
18	-.008	-.000	.001	-.002	-.002	-.006	-.002	.004	.001	.004
19	-.005	-.000	-.001	-.002	-.002	-.006	-.000	.001	.002	.002
20	-.004	-.002	.001	-.001	-.004	-.007	-.002	.004	-.001	.002
21	-.003	.000	.001	.001	.000	-.004	-.002	.005	.003	.002
22	.002	.006	.005	.015	.010	.008	.001	.010	.005	.004
23	.029	.042	.027	.053	.045	.048	.013	.043	.016	.013
24	.130	.109	.073	.061	.045	.063	.033	.072	.065	.050
25	.073	.082	.059	.069	.066	.065	.050	.101	.087	.061
26	.071	.098	.084	.067	.066	.048	.045	.066	.058	.071
27	.084	.074	.083	.064	.073	.065	.052	.053	.074	.074
28	.095	.063	.085	.066	.071	.070	.073	.054	.068	.077
29	.085	.077	.079	.066	.072	.084	.075	.051	.080	.075
30	.094	.074	.078	.073	.078	.081	.076	.065	.086	.069
31	.088	.069	.076	.071	.079	.080	.073	.071	.083	.067
32	.090	.068	.075	.075	.076	.085	.071	.075	.078	.067
33	.084	.070	.078	.070	.076	.082	.073	.075	.080	.069
34	.081	.072	.072	.068	.076	.083	.070	.077	.079	.069
35	.083	.073	.073	.071	.074	.078	.071	.076	.077	.067
36	.082	.076	.072	.074	.075	.077	.075	.081	.078	.065
37	.082	.073	.071	.073	.074	.078	.077	.078	.076	.064
38	.081	.071	.071	.075	.078	.075	.074	.074	.078	.068
39	.080	.073	.071	.076	.076	.071	.072	.073	.079	.068
40	.078	.075	.072	.083	.081	.076	.074	.075	.080	.071
41	.078	.075	.070	.081	.077	.076	.072	.073	.080	.070
42	.078	.074	.073	.075	.079	.077	.072	.073	.078	.070
43	.079	.073	.072	.073	.076	.082	.071	.072	.076	.073
44	.079	.077	.072	.071	.076	.079	.070	.075	.073	.070
45	.079	.076	.073	.076	.075	.082	.068	.074	.074	.070
46	.078	.076	.074	.077	.075	.077	.072	.075	.068	.069
47	.078	.077	.074	.078	.074	.080	.070	.077	.075	.069
48	.077	.077	.073	.080	.074	.081	.070	.076	.074	.073
49	.077	.078	.075	.075	.074	.079	.070	.077	.071	.090
50	.078	.078	.074	.075	.072	.079	.070	.081	.071	.072

b_t (estimates of β_t) produced by 335 with $\pi^{(31)} = \cdot 6$

to a growth change of size 1.5σ at time $t=21$

$$\text{i.e. } \beta_t = \begin{cases} 0 & t<21 \\ 0.075 & t \geq 21 \end{cases}$$

TABLE P.12

t	REALISATION 1	2	3	4	5	6	7	8	9	10
1	.004	-.003	-.001	.006	.000	.002	-.002	-.004	-.003	-.006
2	.000	.003	-.008	-.015	.009	.012	.002	-.001	-.005	-.001
3	.001	.012	.005	.012	.010	-.006	-.001	.021	.009	-.007
4	-.006	.008	.009	.007	.011	-.015	-.000	.017	.006	.002
5	.010	-.008	.014	-.001	.004	-.020	.004	.008	.000	.003
6	.007	-.003	.013	-.011	.003	-.017	.012	.001	.004	.008
7	.007	-.009	.013	-.007	-.005	-.014	.007	.004	-.006	.010
8	.005	-.008	.007	-.016	-.007	-.007	.002	.010	-.008	.003
9	.006	-.004	.008	-.009	-.001	.000	.007	.010	-.006	.003
10	-.000	-.001	.008	-.006	.001	-.003	.005	.004	-.006	.006
11	-.002	-.004	.005	-.007	.001	.005	.006	.005	-.001	.002
12	-.000	-.002	.006	-.004	.001	-.002	.000	.002	-.005	-.001
13	-.003	-.005	.004	-.005	-.001	-.005	.003	.002	-.001	-.001
14	-.001	-.005	.002	-.001	-.003	-.005	-.001	.003	.001	.000
15	-.004	-.005	.002	-.006	.001	-.006	-.002	.004	-.003	.001
16	-.007	-.002	.002	-.005	-.000	-.006	-.004	.005	-.002	.002
17	-.008	-.001	.003	-.003	.001	-.005	-.003	.006	-.002	.003
18	-.007	-.001	.001	-.003	-.002	-.006	-.002	.004	.001	.004
19	-.005	-.001	-.000	-.002	-.002	-.006	-.000	.002	.002	.002
20	-.004	-.003	.001	-.002	-.003	-.007	-.002	.004	-.001	.002
21	-.003	-.000	.002	.001	.000	-.004	-.001	.005	.003	.002
22	.001	.006	.005	.021	.012	.011	.002	.009	.004	.003
23	.046	.055	.034	.056	.055	.057	.013	.058	.022	.017
24	.139	.118	.081	.060	.044	.066	.037	.073	.074	.055
25	.071	.077	.055	.070	.073	.064	.057	.108	.092	.062
26	.069	.102	.093	.065	.068	.036	.041	.054	.049	.072
27	.087	.059	.085	.061	.077	.071	.051	.040	.076	.073
28	.099	.049	.087	.066	.071	.077	.092	.051	.065	.078
29	.084	.081	.074	.066	.072	.104	.080	.048	.087	.074
30	.097	.075	.075	.076	.082	.084	.078	.078	.093	.065
31	.087	.065	.074	.072	.081	.081	.070	.081	.083	.064
32	.090	.065	.073	.076	.075	.089	.069	.082	.075	.066
33	.083	.071	.079	.070	.076	.082	.073	.078	.079	.069
34	.079	.073	.068	.067	.077	.084	.068	.081	.078	.068
35	.083	.074	.071	.071	.073	.075	.071	.078	.075	.066
36	.082	.078	.071	.075	.075	.075	.077	.084	.077	.063
37	.081	.072	.069	.073	.073	.077	.078	.080	.075	.063
38	.081	.070	.070	.075	.079	.073	.075	.073	.077	.068
39	.080	.073	.070	.077	.076	.069	.072	.072	.079	.067
40	.078	.075	.072	.085	.082	.074	.075	.075	.080	.071
41	.078	.075	.069	.081	.077	.075	.072	.073	.080	.069
42	.078	.074	.073	.076	.079	.076	.072	.073	.077	.070
43	.079	.074	.072	.075	.076	.083	.071	.072	.076	.073
44	.079	.078	.072	.073	.076	.079	.069	.075	.073	.070
45	.079	.076	.073	.077	.075	.081	.067	.074	.074	.070
46	.077	.076	.074	.078	.075	.077	.071	.075	.069	.069
47	.078	.077	.074	.078	.074	.080	.069	.077	.076	.069
48	.077	.077	.073	.079	.074	.080	.069	.076	.074	.073
49	.077	.078	.075	.075	.074	.079	.070	.077	.071	.094
50	.078	.078	.074	.075	.072	.079	.069	.081	.072	.072

b_t (estimates of β_t) produced by 335 with $\pi^{(33)} = .8$ to a growth change of size $1.5\,\sigma$ at time $t=21$

i.e. $\beta_t = \begin{cases} 0 & t < 21 \\ 0.075 & t \geq 21 \end{cases}$

REFERENCES

[1] Anderson, T.W. An introduction to multivariate Statistical analysis, Wiley, New York. (1958)

[2] Akaike, H. (1976) Canonical correlation Analysis of time series and the use of an information criterion. In R.K.Mehra and D.G.Lainiotis (eds.) System identification : Advances and Case Studies, Academic Press, New York.

[3] Box, G.E.P. and Jenkins, G.M. Time Series Analysis, Holden Day. (1970)

[4] ———— and Muller, M.E. A note on the generation of random normal deviates, Ann.Math.Stats., Vol.29, pp.610-611, 1958.

[5] ———— and Jenkins, G.M. (1968) Some recent advances in forecasting and control, Applied Statistics, Vol.17, pp.91.

[6] Brown, R.G. (1963) Smoothing, forecasting and Prediction, Prentice-Hall, Englewood Cliffs, New Jersey.

[7] ———— (1959) Statistical forecasting for Inventory control, McGraw Hill, New York.

[8] Burgin, T.A. and Wild, A.R. (1967) Stock control-experience and usable theory, ORQ 18, 8.

[9] Cantarelis, N. (1976) A Bayesian forecasting system, University of Warwick, SIBS, M.Sc. thesis.

[10] ———— and Johnston, F.R. (1979) On line variance estimation for the steady state Bayesian forecasting model, University of Warwick, SIBS, Working Paper No.34.

[11] Chatfield C. (1974) The analysis of time series : theory and practice, Chapman and Hall.

[12] Durbin, J. (1959) Efficient estimation of parameters in moving average models, Biometrika, Vol.46, p.306.

[13] Fildes, R.A. (1979) Quantitative forecasting, The state of the art, ORQ, Vol.30, No.8.

[14] Green, M. and Harrison, P.J. (1973) Fashion forecasting for a mail order company using a Bayesian approach, ORQ 24.

[15] Godolphin, E.J. and Harrison, P.J. (1975). Equivalence Theorems for Polynomial Projecting Predictors. JRSS, B.37, 205-215.

[16] Harrison, P.J. (1967) Exponential smoothing and short term sales forecasting, Management Science, Vol.13, No.11.

[17] Harrison, P.J. and Stevens, C.F. Bayes forecasting in action : case studies, University of Warwick, Statistics Research Report No.14.

[18] ———————————— (1976) Bayesian forecasting (with discussion) J.R.Stat.Soc.V, 39.

[19] ———————————————— (1971) A Bayesian approach to short-term forecasting, ORQ, 22, No.4.

[20] ———————— (1965) Short-term sales forecasting, J.R.Stat. Soc.C.XIV, 102-139.

[21] ——————— and Scott (1965) A development system for use in short-term sales forecasting investigations, Paper presented at the O.R.Societies Annual Conference, Shrivenham.

[22] Holt, C.C. (1960) Planning production inventories and work force, Prentice Hall, New York.

[23] ————— (1957) Forecasting seasonals and trends by exponentially weighted moving averages, Carnegia Inst.of Technology : ONR Memorandum 52.

[24] Kalman, R.E. (1960) A new approach to linear filtering and prediction problems, Journal of Basic Engineering, D.82, p.35.

[25] —————— and Bucy, R.S. (1961) New results in linear filtering and prediction theory, Journal of Basic Engineering, D.83, p.95.

[26] Khintchine, A. (1932) Sulle succesioni stationarie di eventi, Giorn. 1st Ital.Attuari 3 p.267.

[27] ——————— (1934) Korrelationstheorie der Stationaren Stochastishen Prozesse, Math.Ann., Vol.109, p.604.

[28] Kolmogorov, A. (1933) Giorn.Ist.Ital.Altuari, Vol.4, p.83.

[29] ——————— (1941) Interpolation and extrapolation of stationary random sequences, Bulleting of Academy of Science, USSR, Math.Ser.Vol.5, p.3.

[30] Leonard, T. and Harrison, P.J. Bayesian updating for the steady state Kalman filter, University of Warwick, Statistics Research Paper (1978).

[31] Lewis, C.D. Scientific Inventory Control, London Butterworths. (1970)

/32/ Makridakis, S. (1978) A survey of time series, Int. Statist Rev. 44, 29-70.

/33/ Mann, H.B. and Wold, A. (1943) On the statistical treatment of linear stochastic difference equations.

/34/ Mehra, R.K. (1976) A survey of time series modelling and forecasting methodology, Address to the International Institute of Applied System Analysis, Laxenberg, Austria.

/35/ Newbold, P. The principles of the Box-Jenkins approach, ORQ, Vol.26, No.2, ii.(1975)

/36/ Slutsky, E. (1937) The summation of random causes as the source of cyclic processes, Econometrica, Vol.5, p.105.

/37/ Stevens, C.F. On the variability of demand for families of items, ORQ 25, No.3. (1974)

/38/ U.S.National Industrial Conference Board, Studies in Business Policy. No.88.

/39/ Walker, A.H. (1931) On theperiodicity in series of related terms, Proceedings of the Royal Society of London, A.131, p. 518.

/40/ ———— (1962) Large sample estimation of parameters for AR processes with MA residuals, Parts I, II Biometrika Vol.49, p.117.

/41/ Whittle, P. (1953) The analysis of multiple stationary time series.

/42/ Wiener N. (1949) Time Series, MIT Press, Cambridge, Mass.

/43/ Winters, P.R. (1960) Forecasting sales by exponentially weighted moving averages, Management Science, p.324.

/44/ Wold, H. (1938) A study in the analysis of stationary time series, Almgrist and Wiksell, Stockholm.

/45/ Wright, A. University of Warwick, SIBS, M.Sc. thesis, 1976.

/46/ Yule, G.U. (1927) On the method of investigating periodicities in disturbed series with special reference to Wölfer's Sunspot Numbers, Philosophical transactions, Series A, Vol.226, p.267.

/47/ Trigg, D.W. and Leach, A.G.(1967) Exponential smoothing with an Adaptive Response rate, ORQ. Vol.18, pp.53-59.

/48/ Chatfield, C. (1978). The Holt-Winters forecasting procedure. Appl.Statist.27, No.3, p.264-279.

/49/ Neave, H.R. (1973). On using the Box-Muller transformation with multiplicative congruential pseudo-random number generators. Appl. Stat.22, No.1, p.92.

/50/ Golder E.R. and Settle J.G. (1976). The Box-Muller method for generating pseudo-random Normal deivates. Appl.Statist.25, No.1, p.12.